

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

Docket No. A-58631-4/RFT/DJM

Anticipated Classification of this Application:

Class: Subclass:

Priority Application

Serial No.: 08/485,355

Filed: June 7, 1995

"EXPRESS MAIL" MAILING LABEL

NUMBER EL605084525US

DATE OF DEPOSIT

I HEREBY CERTIFY THAT THIS PAPER OR FEE IS BEING DEPOSITED WITH THE UNITED STATES POSTAL SERVICE "EXPRESS MAIL POST OFFICE TO ADDRESSEE" SERVICE UNDER 37 C.F.R. 1.10 ON THE DATE INDICATED ABOVE AND IS ADDRESSED TO: ASSISTANT COMMISSIONER FOR PATENTS, BOX PATENT APPLICATION FEE, WASHINGTON, DC 20231.

TYPED NAME VINCENT DIAZ

SIGNED Vincent M. Diaz

Assistant Commissioner for Patents
BOX PATENT APPLICATION FEE
Washington, DC 20231

Sir:

This is a request for filing:

- ☐ Original
- ☐ Continuation
- ☒ Divisional
- ☐ Continuation-in-part

application under 37 C.F.R. 1.53(b), in the name of

PETER D. CHRISTIAN, KARL H.H. GORDON, TERY N. HANZLIK

(Names of ALL Applicants)

for INSECT VIRUSES AND THEIR USES IN PROTECTING PLANTS

(Title of Invention)

This ☐ continuation ☒ divisional ☐ continuation-in-part

claims priority to pending application Serial No. 08/485,355, filed on June 7, 1995.

1. (a) ☐ Enclosed is a new application.
- (b) ☐ Enclosed is a continuation-in-part application.
- (c) ☒ Enclosed is a copy of the prior application.
2. (a) ☐ Enclosed is a new Declaration.
- (b) ☒ Enclosed is a copy of the prior Declaration as originally filed.
3. (a) ☐ Enclosed is a Small Entity Affidavit.
- (b) ☐ A Small Entity Affidavit is of record in the prior application.
4. ☒ The filing fee is calculated below:

Claims as filed in the prior application, less any claims canceled by amendment below:

	(Col 1)	(Col 2)	SMALL ENTITY		OTHER THAN SMALL ENTITY	
	NO. FILED	NO. EXTRA	RATE	FEE	RATE	FEE
BASIC FEE				\$355		\$710
TOTAL CLAIMS	10 - 20 =	0	× 9 =	\$	× 18 =	\$0
INDEP CLAIMS	2 - 3 =	0	× 40 =	\$	× 80 =	\$0
MULTIPLE DEPENDENT CLAIM PRESENTED <u> X </u> NO			+135 =	\$	+270 =	\$0
If the difference in Col 1 is less than zero, enter "0" in Col. 2			TOTAL	\$	TOTAL	\$710

5. ☒ The Commissioner is hereby authorized to charge any additional fees which may be required, including extension fees, or credit any overpayment to Deposit Account No. 06-1300 (File No. A-58631-4/RFT/DJM).
6. ☒ Our check in the amount of \$ 710.00 is enclosed.
☐ The filing fee is NOT being submitted with this transmittal letter.
7. ☒ Cancel in this application original claims 1-14 & 25-32 of the prior application before calculating the filing fee. (At least one independent claim must be retained for filing purposes.)
8. ☒ Amend the specification by inserting before the first line the sentence:
--This is a ☐ continuation ☒ divisional ☐ continuation-in-part
of application Serial No. 08/485,355 filed June 7, 1995.--
9. (a) ☒ Fifty-six sheets of informal drawings are enclosed.
(b) ☐ sheets of formal drawings are enclosed.
10. (a) ☐ Priority of application Serial No. filed on in is claimed under 35 U.S.C. 119.
(b) ☐ The certified copy has been filed in prior application Serial No. filed on .
11. ☒ The prior application is assigned of record to Commonwealth Scientific and Industrial Research Organisation and Pacific Seeds Pty. Ltd.

12. ☒ The power of attorney in the prior application is to:

Name: FLEHR HOHBACH TEST ALBRITTON & HERBERT LLP
Address Four Embarcadero Center - Suite 3400
San Francisco, California 94111-4187

- (a) ☒ The power appears in the original papers in the prior application.
(b) ☐ Since the power does not appear in the original papers, a copy of the power in the prior application is enclosed.
(c) ☐ A new power has been executed and is enclosed.
(d) ☐ Address all future communications to:

RICHARD F. TRECARTIN
FLEHR HOHBACH TEST ALBRITTON & HERBERT LLP
Four Embarcadero Center - Suite 3400
San Francisco, California 94111-4187
Tel.: (650) 494-8700
Fax: (650) 494-8771

13. ☐ A preliminary amendment is enclosed. (Claims added by this amendment have been properly numbered consecutively beginning with the number next following the highest numbered original claim in the prior application.)
14. ☒ I hereby verify that the attached papers are a true duplicate of prior application Serial No. 08/485,355 as originally filed on June 7, 1995.

Date: 10/3/00

Signature: 

Richard F. Trecartin
Reg. No. 31,801

Address of Signer:

FLEHR HOHBACH TEST
ALBRITTON & HERBERT LLP
4 Embarcadero Center - Suite 3400
San Francisco, California 94111-4187
Tel.: (650) 494-8700
Fax: (650) 494-8771

- ☒ Attorney or agent of record
☐ Filed under Section 1.34(a)

INSECT VIRUSES AND THEIR USES IN PROTECTING PLANTS

FIELD OF THE INVENTION

The present invention relates to insect viruses useful in control of insect attack on
5 plants. It particularly relates to biological insecticides, especially those comprised of
insect viruses. In particular applications, the invention also provides recombinant
viruses and transgenic plants.

BACKGROUND OF THE INVENTION

10 There is increasing awareness of the desirability of insect pest control by biological
agents. Considerable effort in recent years has been devoted to the identification and
exploitation of DNA viruses with large genomes, especially the baculoviruses. These
viruses generally require extensive genetic manipulation to become effective
insecticides, and the first such modified viruses are only now being evaluated.

15 In contrast, very little effort has been devoted to the study and use of small viruses
with RNA genomes.

Four main groups of small RNA viruses have been isolated from insects. These
20 include members of the Picornaviridae, the Nodaviridae, the Tetraviridae and the
unclassified viruses. Descriptions of these groups can be found in the Atlas of
Invertebrate Viruses (eds J.R. Adams and J. R. Bonami) (CRC Press, Boca Raton,
1991) and Viruses of Invertebrates (ed. E. Kurstak) (Marcel Dekker, New York, 1991).
These disclosures relating to these viruses concern their pathology and biology, not
25 their use in biological control.

Further information regarding small RNA viruses of insects can be found in P.D. Scotti
et al (1981) "The biology and ecology of strains of an insect small RNA virus
complex" *Advances in Virus Research* 26, 117-143. This review describes the insect
30 picornaviruses cricket paralysis virus and Drosophila C virus (diameters estimated at

Nodaviridae (diameter 29-30 nm, 2 RNA components totalling 4.5 kb)

Picornaviridae (diameter 27-30 nm, one RNA component of 7.5 - 8.5 kb)

Nudaurelia β family (now called Tetraviridae) (diameter around 35 nm, either one RNA of 5.5 kb or two totalling 8 kb)

D. Hendry, V. Hodgson, R. Clark and J. Newman (1985) Small RNA viruses co-infecting the pine emperor moth (*Nudaula cytherea capensis*). *Journal of general Virology* **66**, 627-632 described viruses with mean diameters of 40nm and 38nm and one or two RNA components up to 5.5 kb in length.

Most recently, the term insect small RNA viruses has been used by one of the present inventors to cover three main recognised toxic groups: the Picornaviridae, the Tetraviridae and the Nodaviridae (P.Scotti & P.Christian (1994) The promises and potential problems of using small RNA insect viruses for insect control. *Sains Malaysiana* 23, 9-18).

These references illustrate a long standing usage of the term in this field of the term "small RNA virus" for viruses with certain characteristics as listed above. Another important characteristic of these virus groups is that they are not occluded, in contrast to many large viruses like the cytoplasmic polyhedrosis (RNA) viruses or the DNA baculoviruses, granulosis viruses and entomopox viruses. The term would also be applied to viruses not members of the three families listed above, as long as they satisfied the definition of being up to 40nm in size. There are reports of such unclassified viruses (eg in Hendry *et al.* 1985). Moreover, the taxonomic status of some members of the Tetraviridae still requires clarification and it might even be possible for this family to be split, with HaSV and other members with two RNA components in their genome being separated from those with only one component, like the type member *Nudaurelia* β virus, which has not yet been sequenced. The above definition of "small RNA virus" would still cover all members of such virus families.

SUMMARY OF THE INVENTION

In a first aspect of the present invention there is provided an isolated small RNA virus wherein the virus is up to 40nm in size, is not occluded and infects insect species including *Heliothis* species.

In one particular embodiment, the present invention provides an isolated preparation of *Heliothis armigera* stunt virus referred to as "HaSV" herein.

In a further aspect of the present invention there is provided an isolated nucleic acid molecule comprising a nucleic acid sequence hybridizable with RNA 1 (SEQ ID No: 39) or RNA 2 (SEQ ID No: 47) described herein under low stringency conditions.

5

10

In a yet further aspect the invention provides a recombinant insect virus vector incorporating all or a part of the isolated virus of the present invention.

In a still further aspect of the present invention there is provided a method of controlling insect attack in a plant comprising genetically manipulating said plant so that it is capable of producing HaSV or mutants, derivatives or variants thereof or an insecticidally effective portion of HaSV, mutants, variants or derivatives thereof such that insects feeding on the plants are deleteriously effected. The present invention also provides a transgenic plant so manipulated.

BRIEF DESCRIPTION OF FIGURES

10

Figure 1a is a restriction map of RNA 1 (SEQ ID No. 39) clones.

Figure 1b is a restriction map of RNA 2 (SEQ ID No. 47) clones,

15

Figure 2 is the complete sequence of RNA 1 (SEQ ID No. 39) and of major encoded polypeptide.

Figure 3a is the complete sequence of RNA 2 (SEQ ID No. 47) in the authentic version, and its encoded polypeptides.

20

Figure 3b is the sequence of RNA 2 variant (a 5C version) (SEQ ID No. 51) and its major encoded polypeptide(s).

Figure 4 is bioassay data showing HaSV induced stunting of larvae.

25

Figure 5 is a map of Vector plasmid pT7T2b and PT7T2c.

Figure 6 is a schematic representation of the proteins encoded by RNA 1 (SEQ ID No. 39) and RNA 2 (SEQ ID No. 47).

30

Figure 7 is a schematic representation of the proteins expressed by RNA 2 (SEQ ID No. 47) in bacteria DNA

5 cloning are indicated. (Note that P17 is followed by
 BgIII site, whose
 cohesive ends are compatible with those of BamHI).

Figure 8 illustrates the 3'-terminal secondary structure of HaSV RNAs. The tRNA-like structures at the 3' ends of RNAs 1 and 2 (SEQ ID No. 39 & 47) are shown. Residues in bold are common to both sequences.

Figure 9 Expression strategies for HaSV cDNAs in insect cells. The upper

part of the figure shows the genome organization of RNAs 1 and 2 (SEQ ID Nos. 39 & 47). The lower part shows insertion of cDNAs corresponding to these RNAs into a plasmid vector, between heat shock protein (HSP70) promoter of *Drosophila* and a suitable polyadenylation (pA) signal. The

HSP promoter was obtained by PCR using suitable primers, with a BamHI site inserted by PCR immediately upstream of the start of the transcription, giving the following sequence: GGATCCACAGnnn (SEQ ID No. 1), where the underlined residue is the transcription start site for either RNA. The cDNAs are terminated by Clal sites, allowing direct linkage to ribozyme sequences as described in the text.

Figure 10 Ribozymes to yield correct 3' ends. The sequences of ribozymes inserted as short cDNA fragments into HaSV cDNA clones are shown. The ribozyme fragments were assembled and cloned as described in the text. Designed self-cleavage points are indicated by bold arrows.

Figure 11 Immunoblots to map epitopes on HaSV. A.

Detected with HaSV antiserum. Lane 1: pTP70delSP; lane 2: pTP70; lane 3: pTP17; lane 4:

control; lane 5: pTP70delN; lane 6: pTP70; lane 7: pTP71; lane 8: HaSV virions; lane 9: molecular weight markers. B. Detected with HaSV

antiserum. Lane 1: pTP70delN; lane 2: pTP70 delSPN; lane 3: pTP70. C. Detected with an antiserum to the Bt toxin (CryIA(c)). Lane 1: pTP70; lane 2: HaSV virions; lane 3: control extract.

Figure 12 New field isolates of HaSV. The genomic organization of RNA 2

is shown at the top of the Figure. PCR using appropriate primers with BamHI restriction sites and in some cases altered context sequences of the

006007 " 6597360

AUG initiating translation of the P17 (SEQ ID No. 48) or P71 (SEQ ID No. 50) genes were used to make fragments for cloning into the BamHI sites of the expression vectors. Constructs 17E71 (SEQ ID No. 35) and P71 (SEQ ID No. 50) have altered context sequences of the AUG initiating translation of the P17 (SEQ ID No. 48) and P71 (SEQ ID No. 50) genes respectively; these alterations correspond to the context derived from the JHE gene (see text). All context sequences are given on the right of the Figure. R2 is a clone of the complete RNA sequence as a BamHI fragment in the vector.

Figure 13 Maps of the expression constructs in baculovirus vectors.

Figure 14a to e Various strategies utilizing the present invention.

Figure 15 Expression of RNAs 1 and 2 (SEQ ID Nos. 39 and 47) from baculovirus vectors. The full length cDNA clone of HaSV RNA 1 or 2 (SEQ ID Nos. 39 & 47) was inserted as a BamHI fragment into the baculoexpression vectors. PCR. was used to add BamHI sites immediately adjacent to the 5' and 3' termini of the RNA 1 sequence; sequences of the primers are given in the text. Constructs R1RZ and R2RZ carry cis-acting ribozymes immediately adjacent to the 3' end of the sequence of RNA 1 and 2 (SEQ ID Nos. 39 & 47) respectively.

Figure 16 Expression strategies for HaSV cDNAs in plant cells. The upper part of the Figure shows the

5

C

DETAILED DESCRIPTION OF PREFERRED EMBODIMENTS

A first aspect of the invention contemplates use of small RNA viruses for biological control of insects. In particular, in accordance with the first aspect of this invention there is provided an isolated small RNA virus, particularly *H. armigera* stunt virus or mutants, variants or derivatives thereof capable of infecting insects, in particular the insect species such as *Helicoverpa armigera*. The small RNA virus isolate of the instant invention is insecticidal and in particular stunts the growth of insect larvae, for example *Helicoverpa armigera* larvae and inhibits or prevents development into the adult stage.

The small RNA viruses of the instant invention have insecticidal, anti-feeding, gut-binding or any synergistic property or other activity useful for insect control.

In particular, *Helicoverpa armigera* stunt virus (HaSV) particles are isometric and approximately 36 nm in diameter with a buoyant density on CsCl gradients of 1.36g/ml. The virus is composed of two major capsid proteins of approximately 64 and 7 KDa in size as determined on SDS-PAGE. The HaSV genome is much larger than the largest known nodavirus (another class of RNA viruses) and comprises two ss (+) RNA molecules of approximately 5.3 and 2.4 kb. The genome appears to lack a blockage of unknown structure at the 3' termini that is found in Nodaviridae. The HaSV genome however shares a capped structure and non-polyadenylation with Nodaviridae. HaSV differs significantly from Nodaviridae and Nudaurelia w virus in terms of its immunological properties. In particular the large capsid protein has different antigenic determinants. Other properties of HaSV are described in the Examples.

The host range of HaSV includes Lepidopterans such as from the subfamily Heliothinae. Species known to be hosts are *Helicoverpa (Heliothis) armigera*, *H. punctigera*, *H. zea*, *Heliothis virescens* and other such noctuides as *Spodoptera exigua*. *H. armigera* which is known by the common names corn ear worm, cotton ball worm, tomato grub and tobacco bud worm is a pest of economic significance in most countries. *H. punctigera*, the native bud worm, is a pest of the great economic significance in Australia. Members of the Heliothinae, which include *Helicoverpa* and *Heliothis*, and especially *H. armigera* are among the most important and widespread pests in the world. In the US *Heliothis virescens* and *Helicoverpa zea* are particularly important pests.

The first aspect of the invention provides an isolated small RNA virus capable of infecting insects including *Heliothis* species. In a particularly preferred form the invention relates to mutants, variants and derivatives of HaSV. The terms "mutant", "variant" and "derivative" include all naturally occurring and artificially created viruses or viral components which differ from the HaSV isolate as herein described in nucleotide content or sequence, amino acid content or sequence, immunological reactivity, non-glycosylation or glycosylation pattern and/or infectivity but generally retain insecticidal activity. Specifically the terms "mutant", "variant" and "derivative" of HaSV covers small RNA viruses which have one or more functional characteristic of HaSV described herein. Examples of mutants, variants or derivatives of HaSV include small RNA viruses that have different nucleic or amino acid sequences from HaSV but retain one or more functional features of HaSV. These may include strains with genetically silent substitutions, strains carrying replication and encapsidation sequences and signals that are functionally related to HaSV, or strains that carry functionally related protein domains.

In a preferred aspect the invention relates to mutants, variants or derivatives of HaSV which encode replication or encapsidation sequences, structures or signals with 60%, preferably 70%, more preferably 80%, still more preferably 90% and even more preferably 95% nucleotide sequence-identity to the nucleotide sequences HaSV.

5

10

RNA 2 (SEQ ID No: 47) (in the capsid protein)

- 15

20

25

30

nucleotide sequence under low, preferably medium and most preferably high stringency conditions (Sambrook J, Fritsch, E.F. & Maniatis T. (1989) Molecular Cloning: A Laboratory Manual, 2nd Edition, Cold Spring Harbour Laboratories Press) and to nucleotide sequences encoding functionally equivalent sequences. In still a more preferred embodiment the invention comprises the nucleotide sequences of genome components 1 and 2 (SEQ ID Nos: 39 and 47) as represented by Figures 1 and 2 hereinafter or parts thereof, or mutants, variants, or derivatives thereof. The terms "mutants", "variants" or "derivatives" of nucleotide genome components 1 and 2 (SEQ ID Nos: 39 and 47) has the same meaning, when applied to nucleotide sequences as that given above and includes parts of genome components 1 and 2 (SEQ ID Nos: 39 and 47).

The second aspect of the invention also relates to nucleotide signals, sequences or structures which enable the nucleic acid on which they are present to be replicated by HaSV replicase. Furthermore the invention relates to the nucleotide signals, sequences or structures which enable nucleic acids on which they are present to be encapsidated.

In a particularly preferred embodiment of the second aspect, the invention comprises nucleotide sequences which are mutants of the capsid gene having the following sequences:

ATG GGC GAT GCC GGC GTC GCGT TCA CAG (SEQ ID No: 2)

ATG GAG GAT GCT GGA GTG GCG TCA CAG (SEQ ID No: 3)

ATG AGC GAG GCC GGC GTC GCG TCA CAG (SEQ ID No: 4)

In a preferred aspect the invention relates to nucleotide sequences of HaSV encoding insecticidal activity including the capsid protein gene and P17 (SEQ ID No: 48) and mutants, variants and derivatives thereof.

In another preferred aspect the invention comprises nucleotide sequences including the following ribozyme oligonucleotides:

5' CCATCGATGCCGGACTGGTATCCCGAGGGAC (called "5'HVR2C1a" herein)
(SEQ ID No: 6)

10 5' GCTCTAGATCCATTCGCCATCCGAAGATGCCCATCCGGC (called
"RZHDV2" herein) (SEQ ID No: 8)

15 5' GCTCTAGACCAGGTAATATACCACAACGTGTGTTTCTCT (called "RZHC2"
herein) (SEQ ID No: 10)

Ribozyme sequences are useful for obtaining translation, replication and encapsidation of the transcript. It is therefore desirable to cleave the transcript downstream of its t-RNA-like structure or poly A tail prior to translation, replication or encapsidation of the transcript.

The present invention also further extends to oligonucleotide primers for the above sequences, antisense sequences and nucleotide probes for the above sequences and homologues and analogues of said primers, antisense sequences and probes. Such primers and probes are useful in the identification, isolation and/or cloning of genes encoding insecticidally effective proteins or proteins required for viral activity, from HaSV or another virus (whether related or unrelated) carrying a similar gene or similar RNA sequence. They are also useful in screening for HaSV or other viruses in the

field or in identifying HaSV or other viruses in insects, especially in order to identify related viruses capable of causing pathogenicity similar to HaSV.

Any pair of oligonucleotide primers derived from either RNA 1 or RNA 2 (SEQ ID Nos: 39 and 47) and located between ca 300 and 1500 bp apart can be used as primers. The following pairs of primer sequences exemplify particularly preferred embodiments of the present invention: Specifically for RNA 1 (SEQ ID No: 39):

1. HVR1B5' (SEQ ID No: 38) (described below) and the primer complementary to nucleotides 1192-1212 of Figure 1.
2. The primer corresponding to nucleotides 4084 and 4100 of Fig. 13 and the primer HVR13p (SEQ ID No: 12) described below

Specifically for RNA 2 (SEQ ID No: 47):

1. The primer corresponding to nucleotides 459 to 476 of Fig. 2 and the primer complementary to nucleotides 1653 to 1669 of Fig. 2 (this would include the central variable domain)
 2. R2cdha5 and the primer complementary to nucleotides 1156 to 1172 of Fig. 2
 3. The primer corresponding to nucleotides 1178 to 1194 and the primer complementary to nucleotides 2072 to 2091 (of Fig. 2).
- Other combinations giving shorter fragments are also possible.

Further preferred primers include:

5' GGGGGGAATTCATTTAGGTGACACTATAGTTCTGCCTCCCCGGAC (called "HvR1SP5p" herein) (SEQ ID No: 11)

5' GGGGGGATCCTGTTATCCCAGGGGGGC (called "HvR13p" herein) (SEQ ID No: 12)

5' CCGGAAGCTTGTTTTTCTTTCTTTACCA (called "Hr2cdna5" herein) (SEQ ID No: 13)

TCAGCAGGTGGCATAGG (called "HvR23p" herein) (SEQ ID No: 14)

5 C (called "HVPET65N" herein) (SEQ ID No: 15)

TCAC (called "HVPET63N" herein) (SEQ ID No: 16)

10 GGAGATCTACATATGGGAGATGCTGGAGTG (called "HVPET64N" herein)
(SEQ ID No: 17)

GTAGCGAACGTCGAGAA (called "HVRNA2F3" herein) (SEQ ID No. 18)

15 GGGGGATCCTCAGTTGTCAGTGGCGGGGTAG (called "HVP65C" herein) (SEQ
ID No: 19)

GGGGATCCCTAATTGGCACGAGCGGCGC (called "HVP6C2" herein) (SEQ ID No: 20)

20
AATTACATATGGCGGCCGCCGTTTCTGCC (called "HVP6MA" herein)
(SEQ ID No: 21)

AATTACATATGTTTCGCGGCCCGCCGTTTCT (called "HVP6MF" herein)
(SEQ ID No: 22)

5 The invention also relates to vectors encoding the nucleotide sequence described above and to host cells including the same. Preferably these vectors are capable of expression in animal, plant or bacterial cell or are capable of transferring the sequences of the present invention to the genome of other organisms such as plants. More preferably they are capable of expression in insect and crop plant cells.

10 In a preferred aspect the invention relates to the vectors pDHVR1, pDHVR1RZ, pDHVR2, pDHVR2RZ, p17V71, p17E71, pPH, pV71, p17V64, p17E64, pP64, pV64, pBacHVR1, pBacHVR1RZ, pBacHVR2, pBacHVR2RZ, pHSPR1, pHSPR1RZ, pHSPR2, pHSPR2RZ, pSR1(E3)A, pSR1(E3)B, pSR2A, pSR2B, pSX2P70, pSXR2P70, pSRP2B, pBHVR1B, pBHVR2B, pT7T2P64, pSR2P70, pT7T2P65,
15 pT7T2P70, pT7T2-P71, pBSKSE3, pBSR15, pBSR25p, pSR25, phr236P70, phr235P65, pGemP63N, pGemP64N, pGemP65N, pP64N, pP65H, pTP6MA, pTP6MF, pTP17, pTP17delBB, pP656 or p70G as described hereinafter.

20 In a third aspect the invention relates to polypeptides or proteins encoded by HaSV and to homologues and analogues thereof. This aspect of the invention also relates to derivatives and variants of the polypeptides and proteins of HaSV. Such derivatives and variants include substitutions and/or deletions of one or more amino acids, and amino and carboxy terminal fusions with other polypeptides or proteins. In a preferred aspect the invention relates to the proteins P7, P16, P17 (SEQ ID No: 48), P64, P70
25 (SEQ ID No: 52), P71 (SEQ ID No: 50), P11a (SEQ ID No: 42), P11b (SEQ ID No: 44), P14 (SEQ ID No: 46) and P187 (SEQ ID No: 40) described herein and to homologues and analogues thereof, including fusion proteins particularly of P71 (SEQ ID No: 50) such as P70 (SEQ ID No: 52) described herein. In a most preferred aspect the invention relates to polypeptides or proteins from HaSV which have insecticidal
30 activity themselves or provide target specificity for insecticidal agents. In particular the invention relates to polypeptides or fragments thereof with insect gut binding

specificity, particularly to the variable domains thereof as herein described. In addition, homologues and analogues with said insecticidal activity of the polypeptides and proteins are also included within the scope of the invention. In addition the invention also relates to antibodies (such as monoclonal or polyclonal antibodies or chimeric antibodies including phage antibodies produced in bacteria) specific for said polypeptide and protein sequences. Such antibodies are useful in detecting HaSV and related viruses or the protein products thereof.

In a fourth aspect the invention provides an infectious, recombinant insect virus including a vector, an expressible nucleic acid sequence comprising all of, or a portion of the HaSV genome, including an insecticidally effective portion of the genome and optionally, material derived from another insect virus species or isolate(s).

Insect virus vectors suitable for the invention according to this aspect, include baculoviruses, entomopoxviruses and cytoplasmic polyhedrosis viruses. Most preferably, the insect virus vector is selected from the group comprising the baculovirus genera of nuclear polyhedrosis viruses (NPV's) and granulosis viruses (GV's). In this aspect of the invention the vector acts as a carrier for the HaSV genes encoding insecticidal activity. The recombinant insect virus vector may be grown by either established procedures Shieh, (1989), Vlcek (in press) or any other suitable procedure and the virus disseminated as needed. The insect virus vectors may be those described in copending International application No. PCT/AU92/00413.

The nucleic acid sequence or sequences incorporated into the recombinant vector may be a cDNA, DNA or RNA sequence and may comprise the genome or portion thereof of a DNA or RNA of HaSV or another species. The term "material derived from another insect virus species or isolate" includes any nucleic acid sequence, or protein sequence or parts thereof which are useful in exerting an insecticidal effect when incorporated in the recombinant vector of the invention. Suitable nucleic acid sequences for incorporation into the recombinant vector include insecticidally effective agents such as a neurotoxin from the mite *Pyemotes tritici* (Tomalski, M.D.

& Miller, L.K. Nature 352, 82-85 (1991) a toxin component of the venom of the North African scorpion *Androctonus australis* Maeda, S. et al. Virology 184-777-780 (1991) Stewart, L.M.D. et al., Nature 352, 85-88 (1991), Conotoxins from the venom of *Conus spp.* (Olivera B.M. et al., Science 249, 257-263 (1990); Woodward S.R. et al., EMBO J. 9, 1015-1020 (1990); Olivera B.M. et al., Eur. J. Biochem. 202, 589-595 (1991).

The exogenous nucleic acid sequence may be operably placed into the insect virus vector between a viral or cellular promoter and a polyadenylation signal. Upon infection of an insect cell, the vector virus will cause the production of either infectious virus genomic RNA or infectious encapsidated viral particles.

The promoters may be constitutively expressed or inducible. These include tissue specific promoters, temperature sensitive promoters or promoters which are activated when the insect feeds on a metabolite in the plant that it is desired to protect.

Recombinant insect virus vectors according to the present invention may include nucleic acid sequences comprising all or an infectious or insecticidally effective portion of genome the HaSV and optionally another insect virus species or isolate.

In a particularly preferred embodiment of the present invention there is provided assembled capsids comprising one or more of the capsid proteins of the present invention, or derivatives or variants thereof as contemplated or described herein. These assembled virus capsids are useful as vectors for insecticidal agents. As such the assembled viral capsids may be used to administer insecticidal agents such as various nucleotide sequences with insecticidal activity or various toxins to an insect. Nucleotide sequences in the form of RNA or DNA which can be used include those of the HaSV genome or other insect viruses. Toxins which can be used advantageously include those which are active intracellularly and may also include neurotoxins with an appropriate transportation mechanism to reach the insect neurones.

5

10

15

20

30

mentioned copending application, would allow a new level of environmental safety greatly extending earlier approaches based on baculovirus vectors.

ii) Manipulation of encapsidation signals or sequences essential for replicase binding or production of sub-genomic mRNA's including expression of exogenous insect control factors as RNAs dependent on the virus for replication. This involves determination of RNA sequences and signals important for replication and encapsidation of virus RNAs, such as by analysis of replication of deletion mutants carrying reporter genes in appropriate cells, followed by studies on the transmission of the reporter gene to larvae by feeding of virus. These deletion mutants can be used to carry genes for insect control factors/toxins to larvae after replacing the reporter gene by a suitable toxin gene such as shown in Fig. 12;

iii) using an insect promoter responsive to virus infection and, for example, placing copies of the viral replicase gene under the control of two promoters, one which is constitutive or expressed at early stages of vector infection, and the other being a cellular promoter turned on by the ensuing RNA viral infection. This system would then make more copies of the replicase mRNA available as the amount of its template increased. Such a promoter may be isolated using techniques analogous to enhancer trapping, that is, transforming insect cells with a suitable reporter gene and looking for induction of the reporter upon virus infection of a population of transformed cells.

In a fifth aspect the invention relates to a method of controlling insect attack in plants by genetically manipulating plants to express HaSV or parts thereof which can confer insecticidal activity optionally in combination with other insecticidally effective agents. Such plants are referred to as transgenic plants.

The term "express" should be understood as referring to the process of transcribing the genome or portion thereof into RNA or, alternatively, the process of transcribing the

genome or portion thereof into RNA and then, in turn, translating the RNA into a protein or peptide.

In a sixth aspect the invention relates to the transgenic plants per se as described above. Transgenic plants according to the invention may be prepared for example by introducing a DNA construct including a cDNA or DNA fragment encoding all or a desired infectious portion of HaSV, into the genome of a plant. The cDNA or DNA fragment may, preferably, be operably placed between a plant promoter and a polyadenylation signal. Promoters may cause constitutive or inducible expression of the sequences under their control. Furthermore they may be specific to certain tissues, such as the leaves of a plant where insect attack occurs but not to other parts of the plant such as that used for food. The inducible promoters may be induced by stimuli such as disturbance of wind or insect movement on the plant's tissues, or may be specifically turned on by insect damage to plant tissues. Heat may also be a stimulus for promoter induction such as in spring where temperatures increase and likelihood of insect attack also increases. Other stimuli such as spraying by a chemical (for instances a harmless chemical) may induce the promoter.

20 The cDNA or DNA fragment may encode all or a desired infectious portion of the wild-type, recombinant or otherwise mutated HaSV. For example, deletion mutants could be used which lack segments of the viral genome which are non-essential for replication or perhaps pathogenicity.

25 The nucleotide sequences of the invention can be inserted into a plant genome by already established techniques, for example by an Agrobacterium transfer system or by electroporation.

Plants which may be used in this aspect of the invention include plants of both economic and scientific interest. Such plants may be those in general which need protection against the insect pests discussed herein and in particular include tomato, potato, corn, cotton, field pea and tobacco.

To enhance the efficacy of infectious genomic RNA or viral particles expressed by transgenic plants according to the invention, the DNA construct introduced into the plants' genome may be engineered to include one or more exogenous nucleic acid sequences encoding substances that are deleterious to insects. Such substances include, for example, *Bacillus thuringiensis* d-toxin, insect neurohormones, insecticidal compounds from wasp or scorpion venom or of heterologous origin, or factors designed to attack and kill infected cells in such a way so as to cause pathogenesis in the infected tissue (for example, a ribozyme targeted against an essential cellular function).

DNA constructs may also be provided which include:

- i) mechanisms for regulating pathogen expression (for example, mechanisms which restrict the expression of ribozymes to the insect cells) by tying for example, expression to abundant virus replication, production of minus-strand RNA or sub-genomic mRNA's; and/or
- ii) mechanisms similar to, or analogous to, those described in copending International patent application number PCT/AU92/00413 so as to achieve a limited-spread system (such as control of replication).

Transgenic plants according to the present invention may also be capable of expressing all or an infectious or insecticidal portion of genomes from HaSV and one or more species or isolates of insect viruses.

In a seventh aspect of the invention HaSV, or insecticidally effective parts thereof, or the infectious recombinant virus vectors of the fourth aspect of the present invention may be applied directly to the plant to control insect attack. HaSV or the recombinant virus vectors may be produced either in whole or in part in either whole insects or in culture cells of insects or in bacteria or in yeast or in some other expression system. HaSV or the recombinant virus forms may be applied in a crude form, semi purified or

purified form optionally in admixture with agriculturally acceptable carrier to the crop in need of protection. HaSV may also be applied as a facilitator of infection where existing insect populations already infected with another agent, such as one or more other viruses whereby HaSV is able to act synergistically to bring about an insecticidal effect. Alternatively HaSV and another agent such as one or more viruses may be applied together to plants to control insects feeding thereon.

A deposit of HaSV No. 18.4 was made on August 5th 1992 at the Australian Government Analytical Laboratories. The deposit was given accession No. N92/35575.

EXAMPLE 1

TAXANOMIC, PHYSIOCHEMICAL AND BIOCHEMICAL CHARACTERISATION OF AN INSECT VIRUS: HaSV

Materials and Methods

A Animals and virus production. H. Armigara larvae were raised as described in Teakle R.E. and Jensen J.M. (1985) *Heliothis punctiger* in Singh P and Moore R.F. (eds) Handbook of Insect Rearing Vol 2., Elsevier, Amsterdam pp 313-322. Larvae were infected for virus production by feeding five day old larvae on 10mg pieces of diet to which 0.064 OD₂₆₀ units of HaSV had been applied. After 24 hours the larvae were then transferred to covered 12-well plates (BioScientific, Sydney, Australia) that contained sufficient diet and grown for eight days after which they were collected and frozen at -80°C until further processed. Frozen larvae were weighed to 100g, placed into 200ml of 50mM Tris buffer (pH 7.4), homogenized, and filtered through four layers of muslin. This homogenate was centrifuged in a Sorvall SS-34 rotor at 10,000 x g for 30 minutes whereupon the supernatant was transferred to fresh tubes and recentrifuged in Beckman SW-28 rotor at 100K xg for 3 hours. The resultant band was collected and repelleted in 50 mM pH 7.2 Tris buffer in a Beckman SW-28 tube by centrifugation at 100K xg for 3 hours. The pelleted virus was resuspended overnight in 1ml of buffer at 4°C then layered onto a

discontinuous CsCl gradient containing equal volumes of 60% and 30% CsCl (w/v) in a Beckman SW-41 tube and centrifuged at 12 h at 200 xg. The resultant pellet was suspended in 100ml of buffer and frozen for further use.

5 **B Particle characterization.** Staining with acridine orange was as described in Mayor H.D. and Hill N.O. (1961) *Virology* 14: p264. Buoyant density was estimated in CsCl gradients according to Scotti P.D., Longworth J.F., Plus N, Crozier G. and Reinganum C. (1981) *Advances in Virus Research* 26: 117-143.

10 **C Immunological procedure.** Rabbit anti-sera to HaSV was produced by standard immunological procedures. Rabbit antisera to the Nudaurelia o virus in addition to the virus itself was provided by Don Hendry (Rhodes University, Grahamstown, South Africa). Rabbit antisera to the Nudaurelia b virus was supplied by the late Carl Reinganum (Plant Research Institute, Burnley, Vic, Australia). The immunological relationship to the Nudaurelia w virus was
15 determined by the standard reciprocal double diffusion technique. Immunoblotting was performed according to Towbin H., Staeheln T. and Gordon J. (1979) *PNAS*. Antibodies monospecific for the major 65 kDa capsid protein were prepared by incubating polyclonal antisera with sections of
20 nitrocellulose blotted with the 65 kDa protein. After extensive washing in Tris buffered saline, the bound antibodies were eluted in 50mM citric buffer, pH 8.0 after a 5 minute incubation.

25 **D Protein characterization.** Polyacrylamide gel electrophoresis in the presence of SDS followed the procedure of Laemmli UK 1970 *Nature* 227: 680-685 and was done with 12.5% gels unless otherwise noted with low and high molecular weight standards from BioRad. Staining was done with a colloidal preparation of Coomassie Blue G-250 (Gradipore Ltd, Pyrmont, New South Wales, Australia). Determination of the M_r of the smallest protein was done with a
30 16% gel and standards of 3.4 kDa, 12.5 kDa and 21.5 kDa (Boehringer Mannheim). Glycosylation of the viral proteins was determined by a general

glycan staining procedure with reagents supplied by Boehringer Mannheim; the positive control was fetuin. N-termini of proteins were sequenced using procedures described by Matsudaira (1989) Purification of Proteins and Peptides by SDS-PAGE in A Practical Guide to Protein and Peptide Purification for Microsequencing ed Matsudaira P.T. Academic Press, San Diego pp 52-72 on an Applied Biosystems 477A gas phase sequencer.

E Nucleic acid characterization. RNA was removed from capsids by twice vortexing a virus suspension with equal volumes of neutralized phenol then with phenol/chloroform (50:50). RNA was then precipitated from the aqueous phase in the presence of 300 mM sodium acetate and 2.5 volumes of ethanol. Digestions of the HaSV nucleic acid with RNase A and DNase I (Boehringer Mannheim) were done with pBSSK(-) phagemid ssDNA and dsDNA (Stratagene) and RNA controls (BRL). Denaturing agarose gel electrophoresis in the presence of formaldehyde was performed according to Sambrook et al (1989). The state of polyadenylation of the viral RNA was determined by two methods. The first method was to compare the binding of identical amounts (20 mg) of viral RNA and poly(A)-selected RNA from *Helicoverpa virescens* to a 1ml slurry of 5mg of oligo-d(T) cellulose (Pharmacia) in a binding buffer consisting of 20 mM Tris pH 7.8, 500 mM NaCl, 1 mM EDTA and 0.04% SDS. The second method was to observe specific priming of viral RNA and viral RNA polyadenylated with poly(A) polymerase (Pharmacia) with d(T)₁₆A/C/G primers in RNA sequencing reactions using reverse transcriptase (US Biochemical) and a protocol provided by the supplier. The 5' cap structure of the genomic RNA and HaSV was determined by observing the ability of polynucleotide kinase to phosphorylate viral RNA with and without preincubation with tobacco acid pyrophosphatase and alkaline phosphatase (Promega) under conditions described by the supplier.

F ***In vitro* translation of HaSV RNA.** *In vitro* translation of HaSV RNA was performed with lysates of both rabbit reticulocytes and wheat germ (Promega) as directed by the supplier. Reactions were conducted in 10 ml volumes with 1.0 mg of RNA in the presence of five μ Ci 35 S-methionine. The labelled proteins were resolved on 10% and 14% SDS-PAGE gels as described above then visualised by autoradiography of the dried gels. The two viral RNAs were separated by a "freeze and squeeze" method after resolution on nondenaturing low-melting-point agarose gels in TAE (Sambrook, et al. 1989). Briefly, agarose slices containing the RNA were melted at 65° C in a volume of TAE buffer equal to six times the agarose volume. The solution was allowed to gel on ice before freezing at -80° C for 30 minutes. The frozen solution was thawed on ice then centrifuged at 14,500xg for 10 minutes after which the supernatant was withdrawn and precipitated by the addition of ethanol.

G **Bioassay of virus-induced pathogenesis**

Known amounts of virus isolate, as shown in Figure 3, were fed to larvae at the growth stages indicated by admixture to standard diet. At the time points shown, the larvae were weighed and the mean and SD calculated. Growth of infected larvae was compared to those of uninfected control populations from the same hatching batch in every experiment.

Results

i) Characteristics and taxonomy of HaSV

The virus particles are isometric and are approximately 36 - 38 nm in diameter. They are composed of two major capsid proteins, of 65 kDa and 6kD is size. The virions contain two single-stranded (+) RNA species of 5.3 kb and 2.4 kb length. The virus bears a similarity in these respects to the Nudaurelia w virus, which has been tentatively regarded as a member of the Tetraviridae; these two viruses differ however,

ii) Particle characterization and serology.

iii) **Protein characterization.**

30

5

10

65 kDa protein:

15

Detailed sequence analysis of the RNA genome carried out in Example 3 showed that RNA 1 (SEQ ID No: 39) encodes a protein of molecular weight 186,980 hereinafter

referred to as P187 (SEQ ID No: 40) and RNA 2 (SEQ ID No: 47) encodes proteins with molecular weight 16, 522 (called P17 (SEQ ID No: 48)) and 70,670 (called P71 (SEQ ID No: 50)). P71 (SEQ ID No: 50) is processed into two proteins of molecular weight 63,378 (called P64) and 7,309 (called P7).

5 iv) Nucleic acid characterization

The extracted nucleic acid from HaSV was readily hydrolysed by RNase A but not by DNase I. Denaturing agarose gel electrophoresis of the extracted RNA genome of HaSV indicated two strands that migrated at 5.5 kb and 2.4 kb. The RNA strands were shown not to have extensive regions of polyadenylation as only 24% of the viral RNA
10 bound to the oligo-d(T) cellulose matrix as opposed to 82% of poly(A)-selected RNA. Further evidence for the non-polyadenylation of the viral genome was provided by the observation that the oligo primer, d(T)₁₆G, gave a clear sequencing ladder using reverse transcriptase only after *in vitro* polyadenylation of the viral strands with poly(A)-polymerase.

15 The demonstration that the strands could be modified with poly(A)-polymerase also showed the lack of any 3' modification. The 5' termini of the viral strands were shown to be capped, most likely with m⁷G(5')ppp(5')G, as they could not be labelled with polynucleotide kinase unless pretreated with tobacco acid pyrophosphatase and
20 alkaline phosphatase.

v) *In vitro* translation.

In vitro translation of the viral RNA yielded different results in the two translation systems used (data not shown). The 5.5 kb RNA translated very poorly in the
25 reticulocyte system whereas it produced in the wheatgerm system more than 20 proteins ranging in size from M_r=195,000 to M_r=12,000. The 2.4 kb viral RNA strand yielded a major protein with an M_r=24,000 in both systems in addition to a minor protein at M_r=70 kDa. A time course of the translation reaction with the 5.5 kb RNA strand showed all labelled proteins were produced at similar rates indicating that the

5 **vi) Presence of another form of HaSV**

Frequently, during purification of HaSV virions, a minor band appeared in varying amounts on the CsCl gradient that had a buoyant density of 1.3 g/ml. On four occasions, when particles from this minor band were used to infect *H. armigera* larvae that were then processed as before for purification of HaSV virions, the HaSV band with a density of 1.296g/ml was again recovered in vast excess to a varying minor amount of the more dense band. No virions of either type were recovered from uninfected control larvae. Proteins extracted from the more dense particles appeared identical to those from the less dense particles when examined by SDS-PAGE and immunoblotting with antibodies specific for the 65 kDa capsid protein of HaSV.

Extraction and examination of the RNA genome with denaturing agarose gel electrophoresis also showed the same 5.5 and 2.4 kb bands. When particles from the more dense band were examined by electron microscopy as before, they appeared to have a larger diameter 45nm but otherwise highly similar to the 38nm particles.

The molar ratio of the two RNA strands was determined by quantitative densitometry of fluorograms of the resolved strands. The ratio derived from an average of four measurements of various loadings on denaturing gels proved to be 1.7:1 (5.5 kb strand: 2.4 kb strand) which is somewhat lower than the expected ratio of 2.3:1 for equimolar amounts of each strand.

The genome of HaSV has major differences that make it distinct from those of the nodaviruses, the only other group of bipartite small RNA viruses pathogenic to animals. Although HaSV shares the characteristic of a bipartite genome with the only animal viruses having such a divided genome, the nodaviridae, it differs in virtually every other aspect from this group. Both segments of its genome are considerably larger than the

0967653 100300

corresponding nodaviral RNAs (Hendry D.A., (1991) Nodaviridae of Invertebrates. in (ed. E. Kurstak) Viruses of Invertebrates. Marcel Dekker, New York, pp. 227-276). However, the division of genetic labour is similar with the larger component carrying the replicase gene and the smaller one encoding the capsid proteins. Direct
 5 comparison of the sequences shows little homology between these viruses, at either RNA or protein level. The Nodaviruses, have the already mentioned unusual 3'blockage (probably a protein), whereas the HaSV RNAs terminate in a distinctive secondary structure resembling a tRNA.

10 **vii) Bioassays of virus isolates on larvae**

The original constructs made to express the capsid proteins (precursor and processed forms) in *E. coli* for bioassay started at the first AUG (nts 284 to 286). Production of full-length, immuno-reactive protein from these was due to these clones being the 5C sequence version with the extra C residue. Bioassays of these proteins have been
 15 difficult due to problems with obtaining suitable *Heliothis* larvae for the tests.

Purified native HaSV was used to conduct bioassays in non-noctuid insect species. The native HaSV was orally administered, the larvae scored for symptoms of infection and growth was measured. Dot blotting for HaSV RNA was also conducted. Based
 20 on these experiments native HaSV does not appear to infect the following larvae.

Species	Order	Family
<i>Galleria mellonella</i>	Lepidoptera	Pyradidae
<i>Tineola bissellia</i>	Lepidoptera	Tineidae
<i>Epiphyas postvittana</i>	Lepidoptera	Tortricidae
25 <i>Lucilia cuprina</i>	Diptera	Calliphoridae
<i>Dacus tyronii</i>	Diptera	Tephritidae
<i>Antitrogus parvulus</i>	Coleoptera	Scarabaediae
<i>Lepidiota picticollis</i>	Coleoptera	Scarabaediae
<i>Sericesthis germinata</i>	Coleoptera	Scarabaediae

The above experiment conducted with the larvae of *Spodoptera exigua* and *S. litura* showed that native HaSV infects these species but not to the same degree as seen in *Heliothis armigera*.

5

EXAMPLE 2

OTHER VIRUS ISOLATES

Materials and Methods

A Virus isolation

10 Apparently infected (*viz* diseased) larvae of *Helicoverpa sp* were collected in February 1993 at Mullaley (NSW), Narrabri (NSW) and Toowoomba (QLD) (Australia).

Referring to Fig. 10 the samples in wells 2A-2D were from parasitised *H. armigera* larvae collected from sorghum at Mullaley; the sample in 6C was collected from sunflower at Toowoomba; the sample in 7D was collected from cotton at the Narrabri
15 Research Station. The latter two larvae may have been either *H. armigera* or *H. punctigera*, which are both easily infected with HaSV.

B Virus RNA Extraction

Larvae collected were ground up and RNA extracted. RNA extraction and purification
20 were as per Example 1.

C Dot-Blot Northern Hybridization

Extracts of viral RNA was analysed by Northern dot-blot hybridisation using a probe made from cloned HaSV sequences derived from 3'-terminal 1000 units of RNA 1 and
25 RNA 2 by random priming in a Boehringer Mannheim kit according to the supplier's instructions were employed. RNA extracts were transferred to Zeta-Probe (BioRad) for probing. Hybridization under high stringency washing conditions were as specified by BioRad. Hybridizations were carried out in the following solution:

1 mM EDTA, 500 mM NaH_2PO_4 , pH 7.2, 7% SDS, at 65°C in a
30 rotating Hybaid hybridization chamber. After completion of hybridization and removal of the solution containing the probe, the

filters were washed twice in 1 mM EDTA, 40 mM NaH_2PO_4 pH 7.2, 5% SDS, at 65°C (1 h each), followed by 2 washes in 1 mM EDTA, 40 mM NaH_2PO_4 , pH 7.2 1% SDS, at 65°C (1 h each), before autoradiography.

5 RESULTS

Referring to Fig. 10, samples 9A, 9B, 10A, 10B and 10C contain HaSV infected positive control lab-raised larvae; 9C-H contain healthy (HaSV-free) negative control lab-raised larvae; All other wells (beginning 1-8) contain extract from field-collected larvae. Numbers 2A-D, 6C and 7D gave positive signals indicating that these isolates
10 are either the same as HaSV or derivatives or variants thereof. Election microscopy employing (-) staining confirmed that the samples which gave positive signals contained abundant icosohedral virus particles of approximately 36nm in size.

The presence of HaSV in larvae which had tested positive in the Northern
15 hybridization dot-blot was confirmed by Western blotting of crude extracts from such infected larvae, using the polyclonal antibody to the HaSV capsid protein. For routine screening of such extracts in order to identify further isolates of HaSV or to confirm the presence of the virus, use of a monoclonal antibody or its equivalent is preferable, in order to achieve (i) higher sensitivity of detection and (ii) greater specificity of
20 detection.

EXAMPLE 3

25 IDENTIFICATION, ISOLATION AND CHARACTERISATION OF INSECT VIRUS GENES

Materials and Methods

A Animals and virus production.

H. armigera larvae were raised as described in Example 1.

30

B Protein characterization

C Nucleic acid characterization

Was conducted as in Example 1.

Was conducted as in Example 1.

The two viral RNAs were separated by a "freeze and squeeze" method after resolution on nondenaturing low melting point agarose gels in TAE (Sambrook, et al, 1989).

E *In vitro* translation of HaSV RNA

F cDNA synthesis and cloning of virus genome

For some clones of RNA2 (SEQ ID No: 47), cDNA was synthesised using the RTase of AMV (Promega) and a specific primer complementary to nucleotide sequence 2285

5

10

15

25

H Subcloning and expression

PCR amplification

The polymerase chain reaction (PCR) was used to obtain sequences covering virus genes in a form suitable for cloning into expression vectors. The reaction was performed with Taq DNA polymerase (Promega) as described by the supplier, in a rapid cycling thermal sequencer manufactured by Corbett Research (Sydney, Australia). A typical reaction involved 1 cycle of 1 min at 90°C, 25 cycles of 95°C (10 sec), 50°C (20 sec), 72°C (1.5 min), followed by one cycle of 72°C for 5 min.

Templates were generally cDNA or cDNA clones derived from HaSV RNAs, made as described below. Primers were as described below for the relevant constructs.

Upon termination of the PCR reaction, the product's ends were made blunt by treatment with E.coli DNA polymerase I (Klenow fragment) at ambient temperature for 15 minutes. After heating at 65° C for 10 minutes, the reaction was cooled on ice and the reaction mix made 1mM in ATP. The product then 5'-phosphorylated using 5 units of T4 polynucleotide kinase at 37° C for 30 minutes. After heating at 65° C for 10 minutes, the product was run on a 1% low-melting agarose gel and purified as described for RNA in section E above.

ligations: Vectors and restriction fragments cut with the enzymes described were run on 1% low-melting-point agarose gels and excised as slices. These slices were then melted at 65° C for 5 minutes, before cooling to 37° C. Fragment and vectors were then ligated in 10ul total volume at 14° C overnight using T4DNA ligase (BRL, Boehringer Mannheim or Promega), in the buffers supplied by the manufacturers.

expression: Expression plasmids containing viral genes (e.g. for the capsid protein) were transformed into *E. coli* strain BL21 (DE3) or HMS174 (DE3) (supplied by Novagen). After growth as specified by the supplier, protein expression was induced by the addition of isopropyl b-D-thiogalactopyranoside (IPTG), at 0.4 nM to the growing culture for a period of 3h. Expressed proteins were analysed by SDS-polyacrylamide gel electrophoresis of bacterial extracts (Laemmli, 1970).

Results

i) Mapping cDNA clones of HaSV

The template for cDNA synthesis was virus RNA which had been polyadenylated in vitro. Oligo(dT) was used as a primer for the Superscript reverse transcriptase (RTase; a modified form of the Moloney murine leukaemia virus (MMLV) RTase, produced by Life Technologies Inc). The cDNA was cloned into vector pBSSK(-) as described earlier. The larger clones were selected for further analysis by restriction mapping and Northern hybridization. All the probes tested hybridized either to RNA 1 or to RNA 2, suggesting that there are no regions of extensive sequence homology between the two RNA's. Furthermore, screening of a number of other clones excluded the theoretical possibility that either RNA band may actually contain more than one species.

ii) RNA 1 (SEQ ID No: 39) clones

Three large RNA 1 (SEQ ID No: 39) clones (B11U, B11O and B35) obtained for the first round of clones were further analysed by restriction mapping and shown to form an overlap spanning over 3 kbp (this was later confirmed by sequencing). The second round of cloning then yielded E3 of 5.3 kbp, representing 99.7% of RNA 1 (SEQ ID No: 39). A complete restriction map of clone E3 showed it to align with that previously determined for three overlapping clones. On the basis of this alignment, the 5' end of the insert in B11U was placed about 300 nucleotides downstream from the 5' end of the RNA.

Once clones covering a contiguous block had been identified, the orientation relative to the RNA was determined.

iii) RNA 2 (SEQ ID No: 47) clones

Three significant cDNA clones were isolated for RNA 2 (SEQ ID No: 47) (Fig. 2).

One, hr236, contains about 88% of RNA 2 (SEQ ID No: 47) (2470 bp total length), and runs from the 3' end to 240 bp from the 5' end. The other clones, hr247 and hr 249

position 34 and terminates at nucleotide 5290 and is thought to encode the RNA-dependent RNA polymerase (replicase)(referred to as P187 (SEQ ID No: 40) in Fig. 1) required for virus replication, since it contains the Gly-Asp-Asp conserved triplet and surrounding sequences identified in these enzymes, which are usually large (over 100 kDa), in addition to further homology with the polymerase encoded by tobacco mosaic virus and other plus-stranded RNA viruses.

Referring to Fig. 1 the sequence is presented as the upper strand of the cDNA sequence. This strand is therefore in the same sense as the viral (positive-sense) RNA.

The sequence of the protein encoded by the major open reading frame, encoding the putative RNA-dependent RNA replicase, is shown, as are those of the small open reading frames at the 3' end, corresponding to the proteins P11a (SEQ ID No: 42), P11b (SEQ ID No: 44) and P14 (SEQ ID No: 46).

Clone E3 was inserted downstream of the SP6 promoter for *in vitro* transcription. As mentioned above, the transcript of this clone can be translated in the wheat germ system to yield the 195 kDa protein observed upon translation of fractionated RNA 1 (SEQ ID No: 39) from the virus. The latter yields more lower molecular weight products, presumably due to being contaminated with nicked and degraded RNA. The products derived from the *in vitro* transcript can therefore be regarded as defining the coding capacity of the complete RNA 1 (SEQ ID No: 39) of HaSV.

vi) Sequence of genome component 2 (see Figure 2)

The 2470 nucleotides encode a protein of molecular weight 71,000 which contains the peptide sequences corresponding to those determined from the two virus capsid proteins. This protein is therefore the precursor of these capsid proteins. The protein is a major product of *in vitro* translation of this RNA obtained either from virus particles or by *in vitro* transcription of a full-length cDNA clone; in addition, another major translation product of apparent molecular weight 24,000 is obtained. This protein is derived from a molecular weight 17,000 reading frame overlapping the slab of the capsid protein gene.

Clones hr236 and hr247 were completely sequenced as the first step in RNA 2 sequencing. These sequences were then extensively compared to that obtained by direct RNA sequencing using AMV reverse transcriptase.

- 5 Comparison of the cloned sequence with that by direct RNA sequencing showed both clones lacked 50 nucleotide present in the RNA (at around nucleotide 1500). The sequence of this stretch was obtained by direct RNA sequencing using the AMV RTase. The MMLV "Superscript" RTase, which was used to make all the cDNA clones, was found to simply by-pass this region in sequencing reactions. These 50
10 nucleotides contain a very stable GC-rich hairpin flanked by a 6 bp direct repeat, and the MMLV RTase skips from the first repeat to the second.

- The sequence of RNA 2 (SEQ ID No: 47) was then completed using plasmids pSR2A and pSR2P70 constructed as described below. The plasmids contain a segment of
15 cDNA derived for the AMV RTase, as well as the sequence corresponding to the 5' 240 nucleotides of RNA 2 (SEQ ID No: 47) which are not present on phr236 (Fig. 2). The sequence of RNA in Fig. 2 is presented as the upper strand of the cDNA sequence. This strand is therefore in the same sense as the viral (positive-sense) RNA. The sequences of the proteins encoded by the major open reading frames, encoding the
20 capsid protein precursor P71 (SEQ ID No: 50), and P17 (SEQ ID No: 48).

- The sequence of RNA 2 (SEQ ID No: 47) encodes a major ORF running from a methionine initiation codon at nucleotides 366 to 368 to a termination codon at nucleotides 2307 to 2309. This protein encoded by this ORF has a theoretical
25 molecular weight of 71,000 (SEQ ID No: 50). This initiation codon is in a good context (AGGatgG), suggesting that it will be well recognized by scanning ribosomes. The size of the product is close to that of the residual putative precursor protein identified in purified virus, and to the size of the *in vitro* translation product obtained from RNA 2 (SEQ ID No: 47).

5

10

15

20

30

size. The 24,000 Mr product was also observed upon translation of an *in vitro* transcript covering only nucleotides 220 to 1200 of RNA 2 (SEQ ID No: 47). This region contains no open reading frame other than those already mentioned and cannot encode a protein longer than 157 amino acids.

5

The protein of Mr 24,000 seen upon *in vitro* translation appears to correspond to P17 (SEQ ID No: 48), with the anomaly in apparent size probably being due to the high content of proline (P), glutamate (E), serine (S) and threonine (T). These amino acids cause the protein run more slowly on a gel thereby giving it an apparent size of Mr

10

The Mr 24,000 protein (hereinafter referred to as P17 (SEQ ID No: 48)) may have a function in modifying or manipulating the growth characteristics or cell cycle of HaSV-infected cells. Although a protein of 16kDa (identified in Example 1) is found

15

in small amounts in the capsid, it does not react with antiserum against the virus particles this is unlikely to correspond to P17 (SEQ ID No: 48), since a preparation of the latter proteins migrates with a molecular weight of 24,000 on SDS gels.

Sequence analysis of the Region from nucleotide 500 to 600 of RNA 2 (SEQ ID No: 47) showed that it has the sequence shown in Fig. 2, as do the plasmids pSR2A, pSR2P70, pSR2B and pSXR2P70. However, plasmids pT7T72P65 and pT7T2P70 have an extra C residue at nucleotide 570. The RNA sequence from which they are derived from is shown in Fig. 2 (the "5C" version). In this sequence the first ATG (nucleotides 283 to 285) is in the same reading frame as most of the capsid protein

20

25

gene. The resultant fusion protein is called "P70" (SEQ ID No: 52) and its carboxyterminal-truncated version (a variant of the native P64) is "P65". In view of these clones it was considered important to resolve whether any virus RNA carrying the extra C residue was present in the viral RNA population first isolated for investigation.

30

Direct sequencing of the virus RNA using reverse transcriptase confirmed that the 4C version lacking the extra residue was the abundant form of the RNA. In order to exclude the possibility of a small amount of the RNA having the extra residue, a sensitive PCR assay was designed. This showed that the extra C residue was not present on any RNA in the viral population, and had been introduced into some clones as a PCR artefact. These clones were however retained and used in bacterial expression experiments (below) because of the high level expression obtained of the P65 and P70 (SEQ ID No: 52) fusion proteins.

10 **vii) Comparison with the sequence of the *Nudaurelia* w capsid gene**

The sequence of most of the RNA2 of the *Nudaurelia* w virus has recently been published by Agrawal D.K. and Johnson J.E. (Virology 190 806-814, 1992). From the published sequence it has been determined that this sequence shows 63% homology to that of HaSV RNA2 (SEQ ID No: 47) at the nucleotide level and 66% at the overall amino acid level. A detailed comparison of the capsid proteins of these two viruses shows the amino-terminal 45 residues to be variable, the next 220 residues to be highly conserved, the next 180 residues to be variable and the c-terminal 200 residues covering the small protein P7 to be highly conserved. A more detailed comparison is discussed below.

The published report did not find a complete reading frame corresponding to the 157 amino acid protein (P17 (SEQ ID No: 48)) gene reported above. The AUG is however present, as is a reading frame - starting upstream of the start of the capsid gene - showing considerable amino acid homology to P17 (SEQ ID No: 48) of HaSV. In vitro translation of purified *Nudaurelia* w virus RNA 2 and a re-examination of the nucleotide sequencing data for this RNA may help to resolve the question of whether the *Nudaurelia* w virus also encodes a protein homologous to the HaSV P17.

More interestingly, antisera against these two viruses, which are similar at a nucleotide sequence level, do not show any cross-reactivity.

viii) Construction of full-length clones

RNA 1 (SEQ ID No: 39)

cDNA clone E3, described above contains all but the 5'-18 nucleotides of RNA 1 (SEQ ID No: 39) and included the complete ORF present on the sequence. The first full-length clone of RNA 1 (SEQ ID No: 39) is therefore based on E3. The 4.9 kbp XbaI-ClaI fragment from clone E3 was recloned into pBSKS(-) (Stratagene) cut with XbaI and ClaI, giving pBSKSE3.

The full-length clone of RNA 1 (SEQ ID No: 39) was completed using PCR. The primer defining the 5' end of the RNA carried an EcoRI site, the promoter for the SP6 RNA polymerase and a sequence corresponding to the 5' 17 nucleotides of RNA 1, as shown in Figure 1. The sequence of this primer was:

HvR1SP5p:

5'-GGGGGGAATTCATTAGGTGACACTATAGTTCTGCCTCCCCGGAC (SEQ ID No: 11) (The G which initiates transcription is underlined)

Using an oligonucleotide complementary to nucleotides 1192 - 1212, a PCR product of 1240 bp was efficiently made. The template was cDNA synthesised using the MMLV RTase and the same oligonucleotide complementary to nucleotides 1192 - 1212 was the primer. Upon termination of the PCR reaction, the product's ends were made blunt and then 5'-phosphorylated as described below. The purified PCR fragment was then cleaved with restriction endonuclease XbaI and the 450 bp subfragment corresponding to the 5' end of RNA 1 (SEQ ID No: 39) cloned into the plasmid pBSSK(-)(Stragene) cut with EcoRV and XbaI, to give pBSR15.

To assemble the full-length of RNA 1 (SEQ ID No: 39), pBSKSE3 (above) was cut with XbaI and ScaI giving fragments of 1.2 kbp and 6.8 kbp. pBSR15 was cut with the same enzymes, giving fragments of 2 and 1.8 kbp. Ligation of the 6.8 kbp fragment for pBSKSE3 and the 1.8 kbp fragment for mpBSR15 yielded pSR1(E3)A. Upon linearization at ClaI and *in vitro* transcription with the SP6 RNA polymerase,

and RNA corresponding to RNA 1 (SEQ ID No: 39), and terminating in a poly(A) stretch of about 50 nucleotides, is obtained.

Since the natural RNA 1 (SEQ ID No: 39) does not have a poly (A) tail, an alternative plasmid was constructed which carries a BamHI restriction site immediately downstream of the 3'end of RNA 1 (SEQ ID No: 39). Again this terminal fragment was made using PCR as above. The sequence of the primer was as follows:

HvR13p: 5'-GGGGGGATCCTGGTATCCCAGGGGCGC (SEQ ID No: 12) (the nucleotide complementary to that which was determined as the 3' one, based on its adjacency to the poly(A) stretch, is underlined; RNA terminating at the BamHI site will have the sequence GCGCCCCCUGGGAUACCaggauc (SEQ ID No: 26)).

The template was clone E3 and an oligonucleotide corresponding to nucleotides 4084 - 4100 was the other primer. The 1220 bp product was blunt-ended, kinased and gel-purified as described above, before cleavage with HindIII. The resulting 420 bp subfragment corresponding to the 3' end of RNA 1 (SEQ ID No: 39) cloned into plasmid pSR1(E3)A cut with ClaI, end-filled with Klenow and then cut with HindIII. The resulting plasmid is pSR1(E3)B. Upon linearization at BamHI and *in vitro* transcription with the SP6 RNA polymerase, and RNA corresponding to RNA 1 (SEQ ID No: 39), and terminating as described immediately above is obtained.

ix) RNA 2 (SEQ ID No: 47)

In constructing the full-length cDNA clone to enable *in vitro* transcription of this RNA hr236 described above was used as a basis. Two separate PCR products, one corresponding to the 5' portion of RNA 2 (SEQ ID No: 47), which is missing from this clone altogether, and another covering the region where clone hr236 lacks the hairpin-forming sequence described above, were required.

The primer defining the 5' end of the RNA carried a HindIII site and a sequence corresponding to the 5' 18 nucleotides of RNA 2 (SEQ ID No: 47), as shown in Figure 2. The sequence of this primer was:

Hr2cdna5: 5'-CCGGAAGCTTGTTTTTCTTTCTTTACCA (SEQ ID No: 13)

(The nucleotide underlined corresponds to that identified as the first nucleotide of RNA 2. (SEQ ID No: 47))

Using an oligonucleotide complementary to nucleotides 1653 - 1669, a PCR product of
 5 1.67 kbp was made. The template was cDNA synthesised using the MMLV RTase
 and an oligonucleotide complementary to the 18 nucleotides at the 3' end of RNA 2
 (SEQ ID No: 47) as the primer. Upon termination of the PCR reaction, the product
 was blunt-ended, kinased and gel-purified as described above, before cleavage with
 PstI. The resulting 1.3 kbp subfragment corresponding to the 5' half of RNA 2 (SEQ
 10 ID No: 47) was cloned into plasmid pBSSK(-) (Stragene) cut with EcoRV and PstI,
 giving plasmid pBSR25p. In order to place this subfragment corresponding to the 5'
 half of RNA 2 (SEQ ID No: 47) downstream of the SP6 promoter for *in vitro*
 transcription, a 1.3 kbp HindIII - BamHI fragment was excised from pBSR25p and
 ligated into HindIII - BamHI cut pGEM-1 (Promega), giving plasmid pSR25.

15 The second PCR product, covering the region where clone hr236 lacks the hairpin-
 forming sequence described above, was synthesised using as primers oligonucleotides
 corresponding to nucleotide sequence 873 to 889 of RNA 2 (SEQ ID No: 47) and to
 the complement of nucleotide sequence 2290 - 2309. Upon termination of the PCR
 20 reaction, the product was blunt-ended, kinased and gel-purified as described above,
 before cleavage with AatII. The resulting 1.1 kbp subfragment covering the required
 region was cloned into plasmid phr236 cut with HindIII, end-filled with Klenow and
 cut with AatII, giving plasmid phr236P70.

25 The two segments were joined covering the first 230 nucleotides of RNA 2 (SEQ ID
 No: 47) together. Plasmid phr236P70 was cut at the SacI site in the vector adjacent to
 the 5' end of the insert and this made blunt-ended using Klenow in the absence of
 dNTPs. After heat-inactivation of the Klenow, the plasmid was cut with EcoRI,
 yielding fragments of 4.5 kbp and 380 bp. Plasmid pSR25 was cut with NheI, blunt-
 30 ended by end-filling with Klenow and cut with EcoRI, yielding fragments of 2.8 kbp,
 900 bp and 750 bp. The 4.5 kbp fragment of phr236P70 and the 900 bp fragment of

To complete the full-length clone of RNA 2 (SEQ ID No: 47), it was necessary to insert a fragment covering the 3' end. As with RNA 1 (SEQ ID No: 39), two versions were made. One, called pSR2A, used the 3' end as present in phr236, together with the poly(A) tail present in this version. The other pSR2B, used a PCR fragment carrying a BamHI site immediately downstream of the 3' nucleotide, as in pSR1(E3)B above. To construct pSR2A, a 350 bp NotI-ClaI fragment was excised from phr236 and cloned into pSR2P70 cut with the same endonucleases. Linearization at the unique ClaI site allows *in vitro* transcription of the complete RNA 2 (SEQ ID No: 47) and a poly(A) tail of about 50 nucleotides in length.

The template used was a plasmid phr236. Upon termination of the PCR reaction, the product was blunt-ended, kinased and gel-purified as described above, before cleavage with NotI. The resulting 400 bp subfragment covering the required region was cloned into plasmid pSR2P70 cut with ClaI, end-filled with Klenow and cut with NotI, giving 25 plasmid pSRP2B. Linearization at the unique BamHI site allows *in vitro* transcription of the complete RNA 2 (SEQ ID No: 47), terminating with the sequence ACCaggtac.

30 This plasmid was made to determine where p24 starts. A 2.1 kbp XhoI-BamHI fragment was cut from clone pSR2P70 and ligated into the vector pGEM-1 (Promega) which had been cut with SalI and BamHI. In vitro transcription of the resulting

plasmid after linearization at the unique BamHI site yielded an RNA covering about 70 nucleotides upstream of the first ATG at nucleotides 283 to 286, plus a short sequence derived from the vector.

- 5 In vitro translation of the RNA from pSXR2P70 yielded both proteins (P70 (SEQ ID No: 52) + P24).

xi) Description of virus-induced pathology

10 The virus induces a rapid anti-feeding effect in *Helicoverpa* larvae as determined by experiments with larvae the results of which are shown in Fig. 3. Fig. 3 shows: A. neonate larvae (less than 24 h old) were fed the designated concentrations of isolated virus (in particles per ml [of diet] added to solid diet). They were weighed on following days and the mean of a statistically significant number (24) of larvae shown. Where necessary, mortality was recorded for the higher concentrations. The vertical
15 axis shows the fold-increase in weight from the hatching weight of 0.1 mg per larvae. This scale therefore also corresponds to weight in units of 0.1 mg (ie 300 is equivalent to 30 mg). B. As for A, but the larvae were 5 days old at the start of the virus feeding. The vertical scale is in mg weight.

20 No weight gain at all was detectable with neonates which had been fed the doses of virus over 10^8 particles per ml (virus added to diet). In addition, 100% mortality was evident after four days at the highest doses. Virus doses as low as 10^6 particles per ml (virus added to diet) still cause significant stunting. The five day old larvae showed a cessation of feeding after 48 hours and significant stunting at 4 dpi, but no mortality at
25 comparable virus doses (Figure 3). Neonates are therefore very sensitive indeed to this virus. Virus particles accumulate specifically in the midgut. This potent anti-feeding effect may be due to the capsid protein or another protein encoded by the virus, or to the effect of any combination of such proteins.

30 xii) Expression of virus-encoded proteins in bacteria.

The vectors

The expression system used initially was derived from the pET-11 system (Novagen). Trimmed down versions of pET-11b and c were constructed and used to compare expression of the capsid proteins. However, due to difficulties experienced with this system substantial modification of the original vectors was carried out in order to achieve much higher yields. These results are described in xiii-b) below.

The initial trimmed-down vectors discussed above were made as follows: pGEM-2 (Promega) which carries T7 promoter adjacent to a poly-linker sequence, but has no sequences corresponding to the lac operon, was cut at the unique XbaI (34) and ScaI (1651) sites, giving fragments of 1.61 and 1.25 kbp. The plasmids pET-11b and c were cut with the same enzymes, giving fragments of 4.77 and 0.91 kbp. The 1.61 kbp fragment of pGEM-2, carrying the c-terminal portion of the ampicillin-resistance gene, the origin of replication and the T7 promoter, was then ligated to the 0.91 kbp fragment of the pET vector, which carries a sequence covering the Shine-Dalgarno sequence, the ATG (in a NdeI site), the terminator for the T7 polymerase and the N-terminal portion of the ampicillin-resistance gene. The resulting plasmids of approximately 2.53 kbp, called pT7T2-b and c, therefore carry a complete T7 transcription unit, which may be used as an expression system in a manner similar to the original pET-11 plasmids, but are repressor-neutral within the cell; they neither titrate away repressor by carrying a binding site, nor do they carry the gene producing the repressor. They were found to grow very well in *E.coli* strains JM109 and BL21 (DE3), and to be very efficient expression vectors. The repressor present in the cells was found to be sufficient to keep the genomic T7 polymerase gene uninduced and therefore the foreign gene unexpressed in the absence of IPTG.

xiii-a) Construction of plasmids for expression of capsid proteins

In this section, all proteins expressed from segments of HaSV RNA 2 (SEQ ID No: 47) are referred to by the size of their gene, as defined in Fig. 4 and in section vi) of this example. The following plasmids were constructed by PCR, using the abovementioned full-length clone of RNA 2 (SEQ ID No: 47), plasmid pSR2A as the template, except where mentioned otherwise.

5

10

20

25

30

HVPET63N

(the underlined sequence corresponds to nucleotides 373 to 390 of the sequence shown in Figure 2; the AflII (CTTAAG) and BglII (AGATCT) sites introduced into the sequence by single nucleotide changes (shown in *italics*) in the oligonucleotide are shown in **bold**).

HVPET64N

GGAGATCTACATATGGGAGATGCTGGAGTG (SEQ ID No: 17)

(the underlined sequence corresponds to nucleotides 366 to 383 of the sequence shown in Figure 2; the BglIII site introduced into the sequence by a single nucleotide change in the oligonucleotide is shown in **bold**).

The PCR products obtained from each combination of one of these primers with the abovementioned one were treated with the Klenow fragment of *E.coli* DNA polymerase, and then with T4 polynucleotide kinase in the presence of 1 mM ATP, before purification by agarose gel electrophoresis as described above. Each product was then cleaved with AatII to yield fragments of 0.95 and 0.4 kbp, and each resulting fragment of about .95 kbp cloned into vector pGEM-2 (Promega) cut with HincII and AatII, giving plasmids pGEMP63N (in which the insert commenced with oligonucleotide HVPET63N (SEQ ID No: 16)), pGEMP64N (in which the insert commenced with oligonucleotide HVPET64N (SEQ ID No: 17)) and pGemP65N (in which the insert commenced with oligonucleotide HVPET65N (SEQ ID No: 15)). The fragment covering portion of the HaSV capsid gene was then excised with enzymes AatII and XbaI.

Two versions of plasmid pGemP65N were made, using different templates as described above. pGemP65N was derived from the sequence of the viral RNA, as in plasmid pSF2A; plasmid pGemP65Nc was derived from the sequence carrying an extra C residue, as shown in Fig. 2 (see "5C version").

In parallel, the carboxy-terminal halves of the major capsid protein variant, whether terminating as for P64 or for P71 (SEQ ID No: 50), were also produced using PCR.

An oligonucleotide primer, HVRNA2F3, with the sequence

GTAGCGAACGTCGAGAA (SEQ ID No: 18) (corresponding to nucleotides 873 to

5 889 of the sequence shown in Figure 2) was used in conjunction with each of the two primers following:

HVP65C

GGGGGATCCTCAGTTGTCAGTGGCGGGGTAG (SEQ ID No: 19)

10 (the underlined sequence is complementary to nucleotides 2072 to 2091 of the sequence shown in Figure 2).

HVP6C2

GGGGATCCCTAATTGGCACGAGCGGCGC (SEQ ID No: 20)

15 (the underlined sequence is complementary to nucleotides 2290 to 2309 of the sequence shown in Figure 2).

The PCR products obtained from each combination of one of these primers with the above mentioned one (HVRNA2F3 (SEQ ID No: 18)) were treated with the Klenow fragment of *E. coli* DNA polymerase, and then with T4 polynucleotide kinase in the presence of 1 mM ATP, before purification by agarose gel electrophoresis as described above. Each product was then cleaved with AatII to yield fragments of 0.9 kbp (in the case of HVP65C (SEQ ID No: 19)) or 1.1 kbp (in the case of HVP6C2 (SEQ ID No: 20)) and 0.4 kbp, and each resulting fragment of about .9 or 1.1 kbp cloned into 25 plasmid phr236 cut with HindIII, treated with Klenow and AatII, giving plasmids phr236P65C and phr236P70 (which has already been described above), respectively. The fragment covering the c-terminus of the capsid protein gene was then excised with enzymes AatII and BamHI.

30 To assemble plasmids for expression in suitable strains of *E. coli*, the excised XbaI-AatII fragments of 0.95 kbp covering the amino-terminal half of the gene and the

excised AatII - BamHI fragments of 0.9 or 1.1 kbp covering the carboxy-terminal half of the gene were simultaneously ligated into the vector pT7T2 cut with XbaI and BamHI. Initial transformation was of *E. coli* strain JM109. Recombinant plasmids carrying the correct insert were then transformed into strain BL21(DE3) for expression as described above.

The plasmid obtained by ligating the aminoterminal fragment commencing with oligonucleotide primer HVPET63N (SEQ ID No: 16) to the c-terminal fragment ending at oligonucleotide primer HVP65C (SEQ ID No: 19) in the expression vector pT7T2b was called pP65G.

In the case of plasmid pP64N, containing an insert from HVPET64N (SEQ ID No: 17) to HVP65C (SEQ ID No: 19), the fragment covering the amino-terminal half of the oligonucleotide was excised by BglII and ScaI from the plasmid pGemP64N and the fragment covering the remainder of the gene was excised with ScaI and EcoRI from plasmid pT7T2-P65. These two fragments were then ligated simultaneously into pP65G which had been cut with BglII and EcoRI.

The resulting construct carrying the complete P71 (SEQ ID No: 50) precursor gene was called pT7T2-P71 and that carrying the P64 form of the gene was called pT7T2-P64. In the case of plasmids derived from pGemP65N and pGemP65Nc, carrying inserts commencing as defined by primer HVPET65N, the expression plasmid derived from pGemP65N which is based on PCR products made using as the template the sequence of the viral RNA, as in plasmid pSR2A, was called pTP17; a truncated form of this plasmid, which expresses P17 (SEQ ID No: 48), was made by cutting at the unique BglII and BamHI sites, removing the intervening fragment (which corresponds to the c-terminal part of the insert) and religating the compatible cohesive ends, to give pTP17delBB. The expression plasmids derived from plasmid pGemP65Nc (which was derived from the sequence carrying an extra C residue, were called pT7T2-P65 (carrying an insert terminating at the primer HVP65C (SEQ ID No: 19)) and pT7T2-P70 (carrying an insert terminating at the primer HVP6C2 (SEQ ID No: 20)).

Expression of P6

Two forms of this protein, which arises through processing of the large capsid protein variant precursor P70 (SEQ ID No: 52) and therefore lacks its own initiation codon, were made. One form (protein MA) replaced the phenylalanine at the start of this protein with methionine, giving it the amino-terminal sequence MAA...; the other carries an additional methionine residue, giving it the amino-terminal sequence MFAA... The oligonucleotides used for PCR-amplified products covering the p6 coding sequence carried a NdeI site (bold) at the ATG codon, for direct ligation into the pET-11 vectors. The primers used were:

HVP6MA: AATTACATATGGCGGCCCGCCGTTTCTGCC (SEQ ID No: 21)

HVP6MF: AATTACATATGTTCGCGGCCCGCCGTTTCT (SEQ ID No: 22)

Each of these primers was used in conjunction with primer HVP6C2 (SEQ ID No: 20) to generate a PCR product of 0.2 kbp. These products were blunt-end ligated into vector pBSSK(-) which had been cut with EcoRV and dephosphorylated. The insert corresponding to the p6 gene was excised with NdeI and BamHI (using the BamHI site in the primer HVP6C2 (SEQ ID No: 20)) and ligated into the expression vector pET-11b, which had been cut with the same enzymes. For expression at higher levels, the insert was transferred to PT7T2 as a XbaI - BamHI fragment, yielding plasmids pTP6MA and pTP6MF.

IPTG induction of bacteria containing plasmids pTP6MA or pTP6MF were used to produce p6 for bioassay.

xiii-b) Expression of viral genes in *E. coli* and bioassay in larvae

Expression of P64

IPTG induction of bacteria containing plasmid pT7T2-P65, which contains an insert running from the location of primer HVPET65N (SEQ ID No: 15) to that of primer HVP65C (SEQ ID No: 19), yielded a protein of molecular weight 68 000. This was 3

000 molecular weight greater than the size of the authentic coat protein, as expected. Expression of pP65G, which contains an insert running from HVPET63N (SEQ ID No: 16) to HVP65C (SEQ ID No: 19), yielded a protein of 65 000 molecular weight.

- 5 The authentic capsid protein (P64) was expressed poorly from plasmid pT7T2-P64. Recloning this insert as a NdeI-BamHI fragment back into the other form of the vector (PT7T2b) did not alter this.

Expression of P70

- 10 IPTG induction of bacteria containing plasmid pT7T2-P70, which contains an insert running from the location of primer HVPET65N (SEQ ID No: 15) to that of primer HVP6C2 (SEQ ID No: 20), yielded a protein of molecular weight 73 000. This was 3 000 molecular weight larger than the size of the precursor of the coat protein, as expected.

- 15 The authentic capsid protein precursor (P71 (SEQ ID No: 50)) was expressed poorly from plasmid pT7T2-P71. Recloning this insert as a NdeI-BamHI fragment back into the other form of the vector (pT7T2b) did not alter this.

- 20 Due to the observation mentioned in vi) above, plasmids designed to express all forms of the capsid proteins from several possible ATG's at the start of the open reading frame were constructed.

- 25 It was found that both authentic P64 and P71 (SEQ ID No: 50) were expressed poorly in bacteria. In contrast, P17 (SEQ ID No: 48) and the forms of the capsid protein commencing at the P17 ATG were expressed very well. The extra C residue present in the latter two constructs resulted in a fusion protein being made from these expression plasmid. The sequence of the fusion proteins can be derived from Fig. 2 by including an extra C at position 570. The fusion caused the first 67 residues of the HaSV capsid protein to be replaced by the first 95 residues of P17 (SEQ ID No: 48). Good
30 expression of the large capsid precursor and protein was achieved, but the size of these

proteins were above 3 kDa larger than the authentic forms. Notwithstanding this the expression products of the vectors containing the 5C variant of RNA 2 (SEQ ID No: 51) are still useful because the resulting product, a P70 (SEQ ID No: 52) variant, is only modified at the NH₂ terminus. Since this terminus is thought to be embedded in the capsid structure and therefore not to participate in the initial interaction with the larval midgut cell, the variant is still useful.

In order to produce constructs which ensure that the expressed proteins possessed the native amino terminus, new plasmids carrying the correct sequence were then cloned into the expression vector (pT7T2). It was found these plasmids to express proteins of the correct size.

The P6 has not yet been expressed from the new constructs. No evidence has been found for processing of P70 to yield the mature proteins in bacteria, nor upon *in vitro* translation of synthetic full-length RNA 2 (SEQ ID No: 47).

The P17 (SEQ ID No: 48) gene has also been cloned into the same vectors for expression and bio-assay. This protein accumulates well in bacteria upon induction, and electron microscopy analysis has shown it form spectacular honeycomb-like structures under the bacterial cell wall, completely surrounding the cell interior (results not shown). The properties of this protein including its amino acid composition and ability to form tube-like structures when expressed in bacteria suggest that it may be an homolog of a gap junction protein. The latter is involved in forming the channels linking the cytoplasms of adjacent epithelial cells in the insect gut. P17 could then play a role in enlarging or forming these channels, thereby enabling cell-to-cell movement of the virus in the insect gut, analogous to the movement or spreading proteins encoded by plant RNA viruses.

In order to ensure that the expressed proteins carried the native amino terminus the correct sequence has also been cloned into the expression vector (pT7T2). The vector had been very slightly modified to that described above to introduce two novel

restriction sited (for AflIII and BgIII) flanking the Shine-Dalgarno sequence. The resulting constructs have been found to be poor producers of the capsid proteins. The complete coding regions (which have been completely checked by re-sequencing) have therefore been recloned into the more satisfactory vectors. Results using these constructs suggest that the amino-terminus of the capsid protein presents inherent difficulties in expression. These difficulties may be imposed by either the nucleotide sequence encoding the amino terminus, or the actual amino acid sequence itself. To discriminate between these possibilities, two types of mutants were made in the sequence encoding the amino terminal 5 residues of the HaSV capsid protein. These amino-terminal mutants are as follows:

HVP71GLY

CCCATATG GGC GAT GCC GGC GTC GCG TCA CAG (SEQ ID No: 28)

Met Gly Asp Ala Gly Val Ala Ser Gln (SEQ ID No: 29)

HVP71SER:

CCCATATG AGC GAG GCC GGC GTC GCG TCA CAG (SEQ ID No: 30)

Met Ser Glu Ala Gly Val Ala Ser Gln (SEQ ID No: 31)

Native HaSV seq:

ATG GGA GAT GCT GGA GTG GCG TCA CAG (SEQ ID No. 32)

Met Gly Asp Ala Gly Val Ala Ser Gln (SEQ ID No: 33)

EXAMPLE 4

EXPRESSION IN BACULOVIRUS VECTORS AND BIOASSAY ON LARVAE

Materials and Methods

A(i) Cloning of HaSV capsid protein gene.

The capsid protein gene was amplified by PCR using the following primers:

5' primers:

HV17E71:

5 HVP71:

(SEQ ID No: 36)

5' GGGGGATCCAGCGACATGAGAGATGCTGGAGTGG

10 (SEQ ID No: 37)

The ATG triplets initiating P17 (SEQ ID No: 48) (in HV17V71 (SEQ ID No: 34) and HV17E71 (SEQ ID No: 35)) or P71 (SEQ ID No: 50) (in HVP71 and HVV71) are underlined)

15 3' primers:

Primers HVP65C (SEQ ID No: 19) and HVP6C2 (SEQ ID No: 20), described in Example 3. Results section Xiiiia, were used. These constructs were made using one of the four 5' primers and HVP6C2 (SEQ ID No: 20). Plasmids constructed from PCR products made using one of the four 5'- primers and HVP65C (SEQ ID No: 19) are called 17V64 (made using 5' primer 17E71 (SEQ ID No: 35)), P64 (made using 5' primer P71 (SEQ ID No: 36)) and V64 (made using 5' primer V71 (SEQ ID No: 37)). These plasmids allow expression of P64.

25 **A(ii) Cloning a full length cDNA of HaSV RNA 1 (SEQ ID No: 39).**

For expression of an RNA transcript corresponding to full length HaSV RNA 1 (SEQ ID No: 39), in insect cells by baculovirus infection or plasmid transfection, PCR was used to generate a fragment of cDNA linking the 5' end of RNA 1 (SEQ ID No: 39) to a Bam HI site.

30 The primers were:

HVR1B5'

5' GGGGGATCCGTTCTGCCTCCCCGGAC (SEQ ID No: 38)

(where the underlined nucleotide represents the start of natural RNA 1 (SEQ ID No: 39)), and an oligonucleotide complementary to nucleotides 1192=1212 of RNA 1 (SEQ ID No: 39).

- 5 The template was plasmid pSR1(E3)B described in Example 3 above.

A segment of the 1240 bp PCR fragment corresponding to the 5' 320 nucleotides of RNA 1 (SEQ ID No: 39) was excised with Bam HI and ASC II and cloned into the Bam HI site of pBSSK(-)[Stratagene] together with the 5 kbp ASCII - Bam HI
10 fragment of pSR1(E3)B, giving plasmid pBHVR1B, which carries the complete cDNA to HaSV RNA 1 (SEQ ID No: 39), flanked by Bam HI sites.

A(iii) Cloning a full length CDNA of HaSV RNA 2 (SEQ ID No: 47).

For expression of an RNA transcript corresponding to full length RNA 2 (SEQ ID No:
15 47) in insect cells by baculovirus infection or plasmid transfection, plasmid pB+NR2B was made by inserting a fragment carrying Hind III and Bam HI sites from the multiple cloning site of vector pBSSK(-) [Stratagene] into plasmid pSR2B described above. The resulting plasmid, called pBHVR2B, carried the cDNA corresponding to full length HaSV RNA 2 (SEQ ID No: 47), flanked by Bam HI sites.

20 **A(iv) Baculovirus transfer plasmids.**

Bam HI fragments of 5.3 and 2.5 kbp corresponding to HaSV RNA's 1 and 2 (SEQ ID Nos: 39 and 47) respectively, were excised from pBHVR1B and pBHVR2B respectively and inserted into the baculovirus transfer vectors described below, which had been linearised with Bam HI.

25

B. Baculovirus Expression of Proteins.

Baculovirus transfer vectors and engineered AcMNPV virus were transfected into *Spodoptera frugiperda* (SF9) cells as described by the supplier (Clontech) and as described in the following references:

C. Western Blotting.

D. Oligonucleotides.

10 HVR1Cla

5' HVR2Cla

15

5' CCATCGATGATCCAGCCTCCTCGCGGCGCCGGATGGGCA (SEQ ID No: 7)

20 5' GCTCTAGATCCATTCGCCATCCGAAGATGCCCATCCGGC (SEQ ID No: 8)

5' CCATCGATTTATGCCGAGAAGGTAACCCAGAGAAACACAC (SEQ ID No: 9)

25

5' GCTCTAGACCAGGTAATATACCACAACGTGTGTTTCTCT (SEQ ID No: 10)

30 A series of recombinant baculoviruses has been constructed, based on the pVL941 transfer vector (PharMingen) or pBakPak8 (Clontech) and the AcMNPV. These are

designed to express the correct forms of the precursor and processed HaSV capsid proteins (P64 and P71 (SEQ ID No: 50)) as well as the smaller capsid protein P6, and P17 (SEQ ID No: 48). In all systems where replicatable RNA encoding the nucleotide sequences of the present invention are to be used, such as eukaryotic systems, in order to get efficient replication, translation or encapsidation of the RNA it is necessary to excise structures downstream of the t-RNA like structure such as the 3' extension or poly A tail on the RNA. In order to carry out such an excision, ribozymes or other suitable mechanisms may be employed. This self cleavage activity of the ribozyme containing transcript should proceed at such a rate that most of the transcript is transported into the cytoplasm of the cell before the regeneration of a replicatable 3' end occurs. Such ribozyme systems are more fully explained in Examples 7 and 9. In the results presented here highly efficient production of P64 and P71 (SEQ ID No: 50) has been achieved. Electron microscopy and density gradient analysis have confirmed that empty particles ("capsoids") are being produced in infected cells that efficiently express the P71 precursor gene. P17 (SEQ ID No: 48) placed in the context of the *H. virescens* juvenile hormone esterase (JHE) gene (Hanzlik, T.N., et al, J. Biol. Chem. 264, 12419-25 (1989)) is produced, but not in large amounts. The latter construct results in a reduction of expression of the capsid protein from the same recombinant, presumably due to a reduction in the number of ribosomes reaching the AUG for the capsid gene.

SF9 cells infected with recombinant baculovirus have been shown to contain large amounts of icosahedral virus particles by electron microscopy (data not shown). These particles contained no RNA, and were empty inside. This observation shows that signals on the viral RNA required for encapsidation of RNA must be located in either the 5' 270 nucleotides or the 3' 170 nucleotides, or both, since these sequences were missing from the RNA transcripts made using recombinant baculovirus. Expression of HaSV proteins was confirmed by Western blotting of total protein extracts from infected insect cells.

In addition, the pAcUW31 vector (Clontech), which carries two promoters, is being used to simultaneously express p6 and p64 as separate proteins.

In order to bioassay the capsid protein produced in baculovirus infected cells, it is first necessary to purify it from the baculovirus expression vector. Preliminary attempts
 5 have made use of density gradients, based on the observation that empty virus particles ("assembled capsids") are in fact produced in infected cells.

As outlined earlier, the HaSV genome or portion thereof is a particularly effective insecticidal agent for insertion into baculovirus vectors. Such a vector is constructed
 10 by insertion of the complete virus genome or portion thereof (preferably the replicase gene) into the baculovirus genome as shown in Fig. 13. Preferably the virus genome or replicase is transcribed from a promoter active constitutively in insect cells or active at early stages upon baculovirus infection. An example of such a promoter is the heat shock promoter described in Example 7. Heat shock promoters are also activated in
 15 stressed cells, for example cells stressed by baculovirus infection. An even more preferable use of such a baculovirus construct is to use the HSP promoter to drive the HaSV replicase and another gene for a toxin (as exemplified elsewhere in the specification) where the RNA expressing the toxin gene is capable of being replicated by the HaSV replicase. Such recombinant baculoviruses carrying the HaSV genome
 20 or portions thereof for expression in larvae at early or other stages of the baculovirus infection cycle are particularly effective biological insecticides.

EXAMPLE 5

25 EFFECT OF HaSV GENES AND THEIR PRODUCTS ON PLANTS

Materials and Methods

A. Electroporation of protoplasts.

Protoplasts of *Nicotiana tabacum*, *N. plumbaginifolia* and *Triticum aestivum* and oats were produced and electroporated with either HaSV or HaSV RNA as described in
 30 Matsunaga et al (1992) J.Gen. Virol. 73: 763-766.

B. Northern blot analysis - RNA extraction from protoplasts after harvest

The protoplasts are subjected to 3 cycles of freezing and thawing, and then an equal volume of 2x extraction buffer (100 mM Tris-HCl, pH 7.5, 25 mM EDTA, 1% SDS, made in DEPC treated water) is added, followed by 1 volume of phenol (equilibrated in 10 mM Tris-HCl pH 8.0) heated to 65°C. The samples are mixed by vortexing and incubated at 65°C for 15 min, vortexing every 5 min. After phase separation by centrifugation at room temperature for 5 min, the aqueous phase is re-extracted with phenol, re separated by centrifugation and re-extracted with chloroform/isoamyl alcohol. To the aqueous phase are then added 0.1 volume of DEPC-treated sodium acetate (pH 5.0) and 2 volumes of ethanol. The RNA is recovered by precipitation at -70°C, followed by centrifugation at 4°C for 15 min. The samples were then analysed by agarose gel electrophoresis as described in example 1.

After blotting to Zeta-Probe membrane (BioRad), the hybridization protocols were as above for Example 2.

C. Total protein from HaSV - electroporated protoplasts.

Protoplasts were analysed by SDS-polyacrylamide gel electrophoresis and Western blotting as described in Example 1.

5 Results

i) Use of complete (replication-competent) RNA virus genome in protoplasts

a) HaSV replication in protoplasts

The nodavirus FHV has previously been shown to replicate in barley protoplasts (Selling H.H., Allison, R. F. and Kaesberg, P. Proc. Natl. Acad. Sci. USA 87,434-8
10 (1990). To determine whether HaSV virus RNA can replicate in plants protoplasts, when introduced by electroporation, experiments using protoplasts from *Nicotiana plumbaginifoli* and wheat have been conducted. (These are all species for which protoplasts are regularly available in the Division of Plant industry). Assays for replication including RNA (Northern) blots using probes derived from cloned
15 fragments of cDNA to RNAs 1 and 2 (SEQ ID Nos: 39 and 47), and Western blots, using the antiserum to purified HaSV particles. Initial experiments showed that both HaSV virus and RNA electroporated into protoplasts of *N. plumbaginifolia* resulted in HaSV replication as studied using and verified by northern blots and ELISA. As a positive control TMV RNA was electroporated and was replication observed.

20

b) Bioassays

Protoplasts into which HaSV RNA had been introduced by electroporation were harvested after 6 or 7 days post electroporation and used in bioassays on neonate larvae by addition to normal diet. The results showed significant stunting of test larvae
25 in comparison to control larvae (see Table 1 below). Protoplasts lacking HaSV RNAs had no effect on the larvae, confirming the result of control experiments. This result confirms that HaSV RNA, when expressed or replicated in plant cells, is able to cause the formation of infectious virus particles able to control insect larvae feeding on the plant material.

30

Northern blotting has been used to confirm that RNA electroporation into protoplasts leads to RNA replication.

Table 1: Results of Bioassay from a typical experiment with Nicotiana and oat

5 protoplasts (oat results are shown in brackets) [see over]

	Treatment	Number	Escapes	Number stunted
1.	diet only	12 (12)	2 (3)	0/10 (0/9)
2.	diet+protoplasts	12 (12)	0 (1)	0/12 (0/11)
10 3.	HaSV+diet	12 (12)	0 (1)	12/12 (11/11)
4.	diet+HaSV/protoplasts	12 (n.d.)	0 (n.d.)	12/12 (n.d.)
5.	diet+RNA/protoplasts	12 (12)	0 (0)	11/12 (10*/12)

* HaSV replication in the larvae was confirmed except for two larvae
15 which were dead. The letters "n.d." mean the experiment was not done.

The above results demonstrate assembly of HaSV particles from electroporated RNA in protoplasts of both monocot and dicot plant species.

20 c) Plasmids to test replication of cloned and engineered forms of HaSV

(1) Plasmids allowing in vitro transcription of HaSV RNAs 1 and 2 (SEQ ID Nos: 39 and 47) for electroporation into protoplasts have already been described above.

(2) Plasmids for transient expression of individual HaSV RNAs (1 or 2) (SEQ ID Nos: 39 and 47) in protoplasts. Full-length cDNAs for the two viral RNAs have been
25 inserted into expression plasmids pDH51 (with the CaMV 35 S promoter. Pietrzak M., et al (9186) Nucl. Acids Res. 14, 5857-68) for dicots and pActI.cas (with the rice actin promoter) for monocots (McElroy et al (1990) The Plant Cell 2: 163-171). As with the vectors for expression in insect cells, these expression plasmids are being modified to include a cis-acting ribozyme for generation of authentic ends. The non-
30 ribozyme plasmids gave no virus replication.

ii) Expression of capsid protein in plants

In view of the present inventors' observation that empty particles ("assembled capsids") are being produced in baculovirus-infected cells that efficiently express the P71 precursor gene, expression of the coding region for the capsid protein in tobacco plants was investigated. The vector chosen for this purpose is based on pDH51 which carries the CaMV 35S promoter and polyadenylation signal. If necessary for improved expression, this vector can be modified by the addition of a translation enhancer sequence from e.g. TMV. Although certain groups have constructed transgenic plants expressing the capsid proteins of plant viruses, there has been only one recent report of assembly of empty capsids in such plants (Bertioli et al., (1991) J. gen. Virol. 72: 1801-9). Bertioli et al point out that the protein-protein interactions in most icosohedral plant RNA viruses may be too weak to allow assembly of such capsids. In addition to the present inventors' observation of empty HaSV capsids, it has been found these capsids are very tough, showing great resilience to e.g. repeated cycles of freezing and thawing, so that it is expected to see assembly of empty HaSV capsids ("assembled capsids") in transgenic plants.

Construction of capsid protein expression plasmid.

Vector used was pDH51; linearised with BamHI and phosphatased.

Insert was PCR product made using following 2 primers:

CAPPLANT:

5' GGGGATCC ACA ATG GGA GAT GCT GGA GTC -3'

(BamHI)

(i.e. A BamHI site followed by plant consensus context for ATG of capsid protein

gene and 15 further nucleotides of this gene - nts 366-383 of HaSV RNA2).

The PCR product was made with VENT polymer (New England Biolabs). After gel purification, it was cut with BamHI and cloned into the vector. Orientation screened with EcoRI to identify insert in same direction as promoter giving plasmid

5

10

Materials & Methods

Was as described in Examples 3 and 4.

15

Was as described in Examples 1 and 3.

20

In vitro transcripts of cloned CDNA of HaSV RNA's was translated in vitro as in Examples 1 and 3.

25

Brush Border Membrane Vesicles were prepared from freshly isolated larvae midguts of *H. Armigera* by the method of M. Wolfersberger et al (1987) Comp. Biochem. Physiol. 86A: 301-308, as modified by S.F. Garczyuski et.al. (1991) Applied Environ. Micro-biol 57: 1816-2820. Brush Border Membrane Vesicles binding assays using invitro labelled protein or ¹²⁵ I-labelled protein were as described in Garczynski et.al. (1991) or in H.M.Horton and Burand, J.P. (1993) J.Virol. 67: 1860-1868.

Results

i) Determination of epitopes on the capsid surface

Comparison of the recently published sequence of the *Nudaurelia* ω virus (NwV) capsid protein with that of HaSV shown that these proteins are closely related and fall into four distinct domains, which are alternatively variable and highly conserved.

These domains are summarised as follows:

Residues:	HaSV	1-49	50-272	273-435	437-647
	NwV	1-46	47-269	270-430	431-645
% identity:		37	81	34	81

Comparison of this observation with the alignment by Agrawal and Johnson (1992) between the NwV and the nodavirus BBV (whose crystal structure is known: Hosur et al (1987) *Proteins: Structure, Function & Genetics* 2: 167-176) showed that the variable region coincided with a region forming the most prominent surface protrusion on the BBV capsid. Both HaSV and NwV carry large insertions at this point relative to BBV, and these insertions are largely different in sequence. Assuming that the alignment by Agrawal and Johnson (1992) is correct, then this means that HaSV and NwV have a more prominent pyramid-like structures as a surface protrusion than do the nodaviruses, and the pyramid-like structures are different. As already noted, there is no immunological cross-reactivity between the two viruses, despite the high degree of identity. There is thus a strong implication of the variable domain as a surface protrusion which functions as the sole antigenic region.

To confirm this a 400 bp NarI fragment spanning the variable region was deleted from the capsid gene in the expression vector. With end-filling of these sites the deletion is in-frame, so that a truncated protein of ca. 57 KDa is produced in bacteria upon induction. This protein was recognized only poorly on Western blots by the antiserum against intact HaSV particles made in rabbits. The central variable domain was recognized well by the antiserum when expressed in isolation from the rest of the capsid gene.

As shown in the table above the region of HaSV capsid protein comprising residues 273-439 shows great divergence from the corresponding region of the NwV capsid protein, compared to its immediate flanking regions. Within this region an especially divergent domain is found from residue 351 to residue 411, which shows only 25% identity to the corresponding region of the NwV capsid protein. This region is flanked by the sequences corresponding to the b-sheet structural features b-E(residues 339-349) and b-F(residues 424-431) of the HaSV capsid protein, based on the alignment the NwV and nodavirus capsid proteins by Agrawal and Johnson (1992), and is therefore likely to form the loop of the most prominent surface protrusion on the HaSV capsid. This is based on comparison and correspondence to the nodavirus capsid protein structure and capsid structure as described by Wery J.-P. and Johnson, J.E. (1989) *Analytical Chemistry* 61, 1341A-1350A and Kaesberg, P., et al. (1990) *J. Mol. Biol.* 214, 423-435. This loop is thought to contain important epitopes. It is significant that this exterior loop on the nodavirus capsid protein is one of the most variable regions when capsid proteins sequences from a number of nodaviruses are compared (Kaesberg et al. 1990).

Finally, the present inventors have observed a significant level of immunological cross-reaction on Western blots, between antisera against the CryIA(c) Bt toxin and HaSV capsid protein, whether obtained from virus or expressed in bacteria. Initial data from the NarI deletion mutant described above suggest that this binding is not to the central variable domain, but to other regions of the capsid protein. The only other region of the proteins which shows extensive sequence variability, the amino terminus, cannot be responsible for the binding, since both authentic capsid protein and the protein with an altered amino terminus expressed in bacteria are recognized by the anti Bt antisera.

ii) In-Vitro binding assays

The full-length clones for *in vitro* translation yielding highly ^{35}S or ^3H labelled proteins were constructed by replacing the bacterial translation interaction signal in the T7

plasmids above by the more active eucaryotic context sequence from the JHE gene. The labelled capsid protein made by *in vitro* translation of the *in vitro* transcripts may be tested for binding to brush border membrane vesicles (BBMV's). Conditions are optimised by testing different procedures. The deletion mutant lacking approximately 125 amino acids in the central region, and containing the variable domain, as well as others derived from it are also tested.

iii) Fusion proteins comprising virus capsid midgut binding domains and other proteins

The idea behind these tests is to fuse the binding domain from the HaSV capsid protein to either large proteins (preferably indigestible, causing protein to aggregate in or on the midgut cells) or toxin domains from other proteins with suitable properties but normally different binding specificities (e.g. Bt). In initial experiments, the gene for the complete capsid protein has been fused to the GUS gene, as has a deletion mutant containing essentially only the central portion of the capsid gene. The resulting fusion proteins are being expressed in bacteria and tested for GUS activity, and makes them sensitive probes for binding experiments on midgut tissue.

iv) Mapping binding sites using Bt/HaSV fusion proteins

Analysis of deletion mutants of the CryIA(c) Bt toxin has identified domains which may be involved in determining the host-specificity of this Bt by acting as receptor-binding sites (Schnepf et al (1990) J. Biol. Chem. 265: 20923-20930; Li et al (1991), Nature 353: 815-21. The present inventors have obtained a clone of this toxin gene. Deletion mutants corresponding to those identified by Schnepf et al are constructed. Segments of the HaSV capsid protein gene can then be inserted into these mutants, the protein expressed in bacteria and their insecticidal function assayed.

EXAMPLE 7

VIRAL GROWTH IN CELL CULTURE

Materials & Methods

A. Cell Lines

The following cultured insect cell lines were tested for infection by HaSV:

Drosophila melanogaster, *Helicoverpa armigera* (ovarian derived), *Heliothis zea* (ovarian derived), *Plutella xylostella*, *Spodoptera frugiperda* (SF9).

All lines were grown under standard conditions. Upon reaching confluence, the culture medium was removed and all mono-layers covered with 1.5 ml of cell culture medium into which HaSV had been diluted; the average multiplicity of infection (M.O.I.) was 10^4 . After adsorption at 26°C for 2h, the inoculum was removed, the cells carefully washed twice with phosphate buffered saline (pH 7.0) and incubation continued with 5 ml of 10% Foetal calf serum in TC199 culture medium (Cyto Systems).

B. Northern Blotting Analysis.

Virus replication in all the above cell lines was confirmed by northern blotting analysis. Total RNA was extracted from infected cells by the method of Chomczynski and Sacchi (1987). Anal. Biochem. 162: 156-159. The cells were lysed in 1 ml of lysis solution (4M guanidinium thiocyanate, 25mM sodium citrate, pH 7, 0.5% sarcosyl, 0.1M 2-mercaptoethanol). In order, 0.1 ml of 2M sodium acetate, pH 4, 1 ml of phenol (0.2M sodium acetate equilibrated), and 0.2 ml of chloroform-isoamyl alcohol mixture (49:1) were added with thorough mixing between reagents. This was then vortexed for 10 s and cooled on ice for 15 min. Tubes were centrifuged in an Eppendorf centrifuge at 14k for 15 min at 4°C for at least 15 min to allow RNA precipitation. RNA was pelleted by centrifugation at 14k for 15 min, washed with 0.6 ml of ice-cold 70% ethanol, pelleted once again (10K, 10 min), air dried at room temperature and resuspended in DEPC (Sigma) treated millipore water. RNA was subject to denaturing agarose gel electrophoresis in the presence of formaldehyde according to Sambrook et.al. (1989). The gel was Northern transferred to a zeta-probe membrane (Biorad) as described by Sambrook et.al. (1989). The probe was prepared by random-priming the 3' sequences of the HaSV genome using DNA and cDNA clones pSHVR15GB and pT7T2p71SR-1 as per manufacturer's instructions (Boehringer-Mannheim). Hybridization was carried out as described for the standard

DNA probe protocol contained within the literature for the zeta-probe membrane (Biorad).

C. Vectors

- 5 Vectors as described below.

Results

10 It has been found that HaSV will replicate in several continuous cell lines, of which the best is the *Spodoptera frugiperda* line SF9. Time course assays by Northern blotting in SF9 cells have shown that RNA 1 (SEQ ID No: 39) replication is clearly detectable within a few hours of infection. RNA 2 (SEQ ID No: 47) is present only in very small amounts early in infection and accumulates much more slowly than RNA 1 (SEQ ID No: 39) does. This observation is consistent with one made earlier in HaSV-infected larvae, where RNA 2 (SEQ ID No: 47) replication was not observed until 3
15 days after infection.

Some apparent replication was also observed in *Drosophila* cells (DL2), but with the difference that more RNA 2 (SEQ ID No: 47) replication was observed at the early time points compared to the lepidopteran cell lines above.

20 Plasmids that express the HaSV genome as RNA transcripts from full length cDNA clones have been constructed and tested. These clones, constructed by PCR and carefully checked, have restriction sites immediately adjacent to the ends of the sequence. Transcription is driven from a specially-re-engineered *Drosophila* HSP70
25 promoter.

i) **Constructs for expression in insect cells**

The constructs are based on vectors carrying the *Drosophila* HSP 70 or actin promoters and suitable polyadenylation signals from *Drosophila* (Corces & Pellicer (1984) *J. Biol. Chem.* 259: 14812-14817) or SV40 (Angelichio et al (1991) *Nucl. Acids. Res.* 18: 5037-5043). Since transcription from such plasmids generates viral RNAs carrying long 3' terminal extensions derived from sequences in the polyadenylation signal fragment, it is necessary to achieve cleavage of the transcript immediately after the 3' sequence of the viral RNA. These plasmids gave no virus replication, presumably because of the 3' terminal extension. The method of choice for obtaining authentic 3' termini is based on introduction of DNA sequences encoding a cis-acting ribozyme into the constructs. With suitable engineering, such a ribozyme will cleave immediately 3' to the viral sequences within the transcript. Suitable ribozymes, based on the hepatitis delta virus (Been M.D., Perrotta, A. T. & Rosenstein, S.P. *Biochemistry* 31, 11843-52 (1992) or the hairpin cassette ribozyme (Altschuler, M., Tritz R. & Hampel, A. *Gene* 122, 85-90 (1992) have been designed (see Example 4). This involves synthesis of overlapping oligonucleotides, which are then annealed and end-filled with the Klenow fragment of DNA polymerase, to create short DNA fragments encoding the desired ribozyme. These fragments carry restriction sites at their termini allowing them to be ligated into plasmids between the viral RNA cDNA (which has a 3' restriction site added by PCR) and the restriction fragment carrying the polyadenylation signal. Ribozyme function has been verified (Example 9).

The *Drosophila* HSP70 promoter was joined to the HaSV RNA 1 sequence as follows. A BamHI restriction site was introduced into the promoter sequence as described on p.5 of this specification. Oligonucleotide HVR1B5P described in Example 8 was used to prime PCR of RNA 1 to yield a cDNA copy of the RNA carrying a BamHI restriction site 5' to the RNA 1 sequence and separated from it by the nucleotides ACA which end the HSP70 promoter just before the start of transcription. This common BamHI site was used to link the HSP70 promoter and the HaSV RNA 1 sequence. The resulting plasmid was completed by adding either the hairpin cassette ribozyme

A similar approach was used to obtain plasmids for RNA 2 i.e. pHSPR2HC and
5 pHSPR2HDV.

The latter approach was used to join the sarcoma virus (RSV) long terminal repeat (LTR) promoter to the HaSV cDNAs for expression in insect cells. The RSV LTR promoter is active in many animal cells (Cullen, B.R. Raymond, K. & Ju, G. (1985) Mol. Cell. Biol. 5,438-447) and also in lepidopteran cell lines (D. Miller personal communication). It was obtained from plasmid pRSVCAT (Gorman, C., Padmanabhan, R. & Howard, B.H., (1983) Science 221, 551-553) as a 495 bp fragment carrying a 5'-XbaI site (added by PCR) and terminating at a blunt end with the sequence AACC, with the underlined residue corresponding to that immediately before the start of transcription. The resulting plasmids, pRSVR1HCLA and pRSVR2HCLA, carry the HaSV RNA 1 and 2 cDNAs, respectively, and are otherwise like pHSPR1HC and pHSPR2HC, respectively. These plasmids carry the SV40 late polyadenylation signal. They allow efficient and precise expression of the HaSV genomic RNAs in insect cells, for example if introduced using a baculovirus vector or by transfection.

EXAMPLE 8

SHEDDING OF INFECTED CELLS

Materials & Methods

5 **A. Confocal Laser Scanning Microscopy. (CLSM)**

CLSM enables the visualisation and analysis of three-dimensional cell and tissue structures at the macro and molecular levels. The Leica CLSM used in this example is based on an MC 68020/68881 VME bus (20MHz) with standard 2Mbyte framestore and 4Mbyte RAM and OS9 operating system with programmes written in C code. It
 10 incorporates a Leica Diaplan research microscope and using X10/0.45, X25/0.75, X40/1.30 and X63/1.30 Fluotar objectives has a claimed optical efficiency better than 90%. The confocal pinhole is software controlled over the range of 20 to 200 μ m. Excitation at 488 and 514 nm is provided by a 2 to 50 mW argon-ion laser.

B. Immunocytochemistry (ICC).

15 For whole mount ICC, tissues were dissected under saline and fixed in fresh 4% formaldehyde in phosphate buffered saline (PBS) for at least 15 mins. After multiple washes in PBS they were permeabilized either by 60 mins incubation in PBT (PLBS with 0.1% Triton X-100 plus 0.2% bovine serum albumin). After 30 mins blocking in PBT+N (5% normal goat serum) tissue was incubated in primary antibody diluted
 20 (1:40) in PBT+N for at least 2 hrs at room temperature then at 4°C overnight. After extensive washing in PBT and 30 mins blocking in PBT+N the FITC conjugated secondary antibody diluted (1:60) in PBT+N was incubated for 2 hrs at room temperature plus overnight at 4°C. After multiple washes in PBT and PBS the tissue was cleared in 70% glycerol and mounted in 0.01%w/v p-phenylenediamine
 25 (Sigma#P1519) dissolved in 70% glycerol. All processing was at room temperature unless otherwise stated.

The inventors' current model for the effect of HaSV involves the detection by the insect midgut of infected cells, their identification as infected and their subsequent shedding in numbers sufficient to cause irreparable damage to the insect midgut. The

10 cause severe disruption of the larval gut cells. Thus these two agents appear to act synergistically in causing gut cell disruption.

above). This established that some midgut cells were sufficiently infected with HaSV to give strong fluorescence signals. Such cells were moreover clearly separating from the surrounding tissue, a sign that they were in the process of being shed.

20 Society for Invertebrate Pathology Abstract #96) although in these cases the effect is
too localised and weak to cause any anti-feeding effect apparently only the small RNA
virus of the tetraviridae which are localised to the gut and cause more-or-less severe
anti-feeding effects in their hosts (Moore, N.F. in Kurstak E. (Ed) (1991) Viruses of
Invertebrates. Marcel Dekker, New York pp277-285) are capable of such an effect to
25 an extent sufficient for pest control.

30 into RNA 1 in place of the N-terminal portion of the replicase gene) can be studied. A correlation between virus replication and cell rejection can be confirmed by

histochemical analysis of the midgut cells of the infected larvae. Thus the cell-shedding phenomenon offers a direct and rapid assay for early events in HaSV-infected gut tissue. Extracts of baculo-vector infected insect cells carrying empty HaSV particles can be fed to larvae directly and the midgut examined by toluidine blue staining and immune-fluorescence at intervals after infection. This will allow direct determination of whether the particles can bind the brush border membranes in intact gut, and whether such binding can induce the massive disruption evident in normally infected larvae. Control experiments using extracts from cells infected with the baculovector alone can be conducted to observe and distinguish effects due to the vector. The immune-fluorescence assay on midgut tissue allows analysis of binding to midgut brushborder membranes. Once determined for wild-type capsid protein expressed from a baculo-vector, deletion or replacement mutants can be inserted into the baculovectors. Suitable cell extracts from these can be used to infect larvae.

EXAMPLE 9

ENGINEERED VIRUS AND USES

Materials & Methods

(as indicated in earlier Examples)

i) Engineered virus as a vector for other toxin genes

This involves placing suitable genes under control of HaSV replication and encapsidation signals. Genes which may be suitable include intracellular insect toxins such as ricin, neurotoxins, gelonin and diphtheria toxins. The toxin gene may be placed in the viral gene such that it is a silent (downstream) cistron on a polycistronic RNA, or in a minus strand orientation, requiring replication by the viral polymerase to be expressed. Standard techniques in molecular biology can be used to engineer these vectors.

A discussion of two recombinant HaSV vectors which have been designed is given below:

for RNA 1 (SEQ ID No: 39):

The reporter gene (or one of the toxin genes mentioned above) is inserted in place of the amino-terminal portion of the putative replicase gene, such that the initiation codon used for the replicase (ie that at nucleotides 37-39 of the sequence) is now used to commence reporter gene translation. The fusion is achieved by the use of artificial NcoI restriction sites common to both sequences.

The short 36 nucleotide 5'-untranslated leader of RNA 1 (SEQ ID No: 39) (shown in upper case) is synthesised as the following sequence:

ggggatccacaGTTCTGCCTCCCCCGGACGGTAAATATAGGGGAACCATG

10 Gtctagagg, (SEQ ID No: 53)

using two overlapping oligonucleotides comprising the first 31 (oligonucleotide HVR1B5P) nucleotides and the complement of the last 40 nucleotides (oligonucleotide HVR1NCO) respectively. These primers are annealed and end-filled by Klenow. The resulting fragment is then cut with BamHI and XbaI (sites underlined) and cloned with plasmid vector pBSISK(-) to give pBSSKR1NCO.

15 The GUS gene carrying a NcoI site at the ATG codon was obtained as a NcoI-SacI fragment from plasmid pRAJ275 (Jefferson, RAJ Plant Mol. Biol. Rep 5, 3387-405 (1987)). This SacI site is located just downstream from the coding sequence for the GUS gene.

20 The 5' leader of HaSV RNA1 is excised as a BamHI-NcoI fragment from the plasmid pBSSKR1NCO, and is ligated together with the NcoI-SacI fragment carrying the GUS gene into plasmid pHSPR1HC or pHSPR1HDV or pDHStuR1HC carrying the full-length cDNA insert of RNA 1 (see above) which has been cut with BamHI and SacI. The resulting plasmid then carries a complete form of RNA 1 (SEQ ID No: 39) but
25 with the amino-terminal portion of the replicase gene substituted by the GUS gene. It is desirable to produce a construct with approximately the same size as RNA 1 (SEQ ID No: 39) for encapsidation purposes.

30 Similar approaches are adopted for RNA 2 (SEQ ID No: 47), with the foreign, reporter or toxin gene fused to the initiation codon of either P17 or P71. In either case the context sequence of the introduced gene is modified to give the necessary expression

The above recombinants have been described specifically as insertions of a reporter gene (GUS). The toxin genes to be inserted are described on page 14 of the specification. These preferably further require a signal peptide sequence added at the amino-terminus of the protein.

10

15

20

25

iv)

30

genome, or of the replicase gene as a tool for the amplification of suitable amplifiable mRNAs (e.g. encoding toxin) or of the capsid protein as a means to deliver insecticidal agents. These strategies are now described in some detail.

5 a) Use of the complete HaSV genome

Fragments of cDNA corresponding to the full-length HaSV genome components RNAs 1 and 2 (SEQ ID Nos: 39 and 47) are placed in a suitable vector for plant transformation under the control of either a constitutive plant promoter (e.g. the CaMV 35S promoter mentioned above) or an inducible promoter or a tissue specific (e.g. leaf-specific) promoter. The cDNAs are followed by a cis-cleaving ribozyme and a suitable plant polyadenylation signal. Transcription and translation of these genes in transgenic plant tissues and cells leads to assembly of fully infectious virus particles to infect and kill feeding larvae.

15 The following experiments were conducted. The plasmids for expression used the CaMV 35S promoter to generate transcripts commencing at the first nucleotide of the HaSV RNAs 1 and 2 (SEQ ID Nos: 39 and 47). The vector pDH51 (M. Pietrzak, R. Shilito, T. Hohn and I. Potrykus (1986). *Nucleic Acids Research* 14, 5857) which carries the CaMV 35S promoter followed by a multiple cloning site and the CaMV
20 polyadenylation fragment was modified to make a suitable vector, pDH51Stu, carrying a StuI site at the immediate 3' end of the CaMV 35S promoter. The promoter thereby terminates in the sequence GAGAGGCCT, with the underlined residue being that at which transcription would start. (Similar vectors have been described by Mori *et al.*, *J. General Virology* 72, 243-246 (1991) and Dessens and Lomonosoff, *ibid* 74, 889-892 (1993).) The StuI site (AGG/CCT) is followed by a BamHI site (GGATCC).

25 Cleavage of this vector with StuI and BamHI generates a vector DNA molecule with one blunt end (from StuI cleavage) and one sticky BamHI end. This allows ligation of cDNA molecules corresponding to the full-length HaSV genomic RNAs, and carrying a blunt end at the 5' end of the full-length cDNA and a BamHI site after the 3'-end of
30 the full-length cDNA.

Suitable cDNA fragments carrying a blunt end corresponding to the 5'-terminal nucleotide of either RNA 1 or 2 (SEQ ID Nos: 39 and 47) were generated using PCR and an oligonucleotide primer corresponding to the 5'-terminal first 18 nucleotides of the sequence of either RNA 1 (SEQ ID No: 39) or RNA 2 (SEQ ID No: 47). The

5 cDNA sequence corresponding to the 3' terminal sequences of either RNA 1 (SEQ ID No. 39) or RNA 2 (SEQ ID No 47) were followed on these DNA fragments by sequences corresponding to one of the ribozymes whose sequences are shown in Fig. 8 and whose construction is described in Example 7. The 3'-terminal sequence corresponding to an XbaI site (TCTAGA) shown in these ribozyme sequences was

10 followed on the suitable DNA fragments by a BamHI site, which upon cleavage with this enzyme yielded a sticky end capable of being ligated into the BamHI end of the vector cleaved as described above. There were therefore a total of four suitable DNA fragments for insertion into the vector:

15 RNA 1 (SEQ ID No: 39) followed by the hairpin cassette (HC) ribozyme
 RNA 1 (SEQ ID No: 39) followed by the hepatitis delta virus (HDV) ribozyme
 RNA 2 (SEQ ID No: 47) followed by the hairpin (HC) ribozyme
 RNA 2 (SEQ ID No: 47) followed by the hepatitis delta virus (HDV) ribozyme.

These four fragments were individually ligated into the vector pDH51Stu cleaved with

20 StuI and BamHI to generate four distinct plasmids as follows:
 pDHStuR1HC
 pDHStuR1HDV
 pDHStuR2HC
 pDHStuR2HDV

25 Transcription from the 35S promoter in these plasmids results in RNAs commencing at the first nucleotide of either the RNA 1 sequence (SEQ ID No: 39) or RNA 2 sequence (SEQ ID No: 47) and terminating in the CaMV polyadenylation fragment. Self-cleavage at the locations shown in Fig. 8 by the cis-acting ribozymes obtained within these transcripts generates RNA molecules with the 3'-termini corresponding to

30 the natural virus termini.

After amplification and purification on CsCl gradients, thirty mg of each of these four plasmids was transfected by electroporation into aliquots of two million *N.*

plumbaginifolia protoplasts (as described in Example 5) either individually or in the combinations listed below:

- 5 pDHStuR1HC + pDHStuR2HC
- pDHStuR1HDV + pDHStuR2HDV

The production of infectious HaSV particles within transfected protoplasts was then demonstrated by bioassay on heliothis larvae. After incubation at 25°C for 3-5 days, the protoplasts were recovered by low speed centrifugation and applied directly to standard heliothis diet as surface contamination for bioassay as described in Example 1. Stunting was only observed when plasmids expressing HaSV RNA 1 (SEQ ID No: 39) and RNA 2 (SEQ ID No: 47) were co-transfected, and then only in the case of those carrying the hairpin ribozyme to generate the viral 3' ends (see Table 2). In contrast, constructs carrying the HDV ribozyme at the 3' end were not infectious. The reasons for this have not been determined. As expected, expression of RNA 1 or 2 (SEQ ID Nos: 39 and 47) alone in protoplasts did not lead to the assembly of infectious particles. Western blot analysis of protoplasts transfected with the RNA 2 (SEQ ID No: 47) constructs did show production of limited amounts of the capsid protein.

Suitable control experiments confirmed that larval stunting was due to HaSV particles generated *de novo* in the protoplasts. As shown in the Table 2, neither the protoplasts alone nor protoplasts mixed with plasmid DNA were capable of initiating stunting.

Table 2

Treatment		No. of larvae	Escapes	No. stunted
1.	diet alone	24	0	0
2.	diet + HaSV	24	0	24
3.	diet + protoplasts	24	0	0

	4.	diet + pDHStuR1HC	24	0	0
	5.	diet + pDHStuR1HDV	24	0	0
	6.	diet + pDHStuR2HC	24	0	0
	7.	diet + pDHStuR2HDV	24	0	0
5	8.	diet + pDHStuR1HC + pDHStuR2HC*	24	0	22
	9.	diet + pDHStuR1HDV + pDHStuR2HDV*	24	0	0
	10.	pDHStuR1HC + pDHStuR2HC	24	0	0
10		(but mixed with protoplasts)			

* these plasmids were co-transfected with pDHVCAPB (see Example 5)

HaSV infection of stunted larvae was confirmed by dot-blotting of RNA using HaSV specific probes. After weighing, larva were sacrificed and total RNA extracted as follows. Each larva was homogenised in the presence of 260 ml deionised water, 24 ml 2M sodium acetate pH 4.0 and 200 ml phenol equilibrated with 2M sodium acetate pH 4.0. After centrifugation at 14 000 rpm for 15 min at 4°C, the supernatant (about 200 ml) was removed and extracted once with an equal volume of chloroform. After centrifugation, the supernatant (about 200 ml) was mixed with 20 ml of sodium acetate and 400 ml of absolute ethanol. The precipitate after centrifugation was vacuum dried and redissolved in 5-10 ml of sterile, DEPC-treated water. For dot-blotting, the RNA was mixed with 70 ml of DEPC-treated water and 30 ml of 10 mM EDTA, 30 mM NaOH. HaSV RNA was determined and quantified by dot blotting (as described in Example 2) using a probe random primed DNA from clones corresponding to the terminal 1000 nucleotides of RNA 1 and 2. All larvae recorded as stunted in the bioassays were found to carry HaSV and give signals comparable to those of the larvae fed purified HaSV particles (Table 2). To confirm that the larvae were infected with HaSV, ten aliquots of protoplasts were electroporated with plasmids pDHStuR1HC + pDHStuR2HC and the protoplasts fed (after incubation) to 150 heliothis larvae. The larvae were allowed to grow for one week, upon which significant stunting was observed in 50% of the larvae, and virus was then purified from these stunted larvae as described in Example 1. Analysis on CsCl gradients showed the production of distinct

the RNA can be encapsidated by HaSV capsid protein and/or replicated by the HaSV replicase in infected insect cells (see Figs. 12a and 12b)

Transgenic plants would contain two different transgenes, making either
 5 unmodified capsid protein precursor or a modified form in which most of the
 carboxyterminal protein P7 is replaced by a suitable insect-specific toxin or one which
 is inactive as part of a fusion protein. (Gelonin or other ribosome-inactivating
 proteins, insect gut toxins or neurotoxins may be suitable here.) Expression from these
 two transgenes would be regulated so that only the required amounts of the modified
 10 and unmodified forms are made in the plant cell, and assembled in such proportions
 into the capsoids. One way to modulate the production of capsotixin fusion proteins is
 to make translation of the carboxyterminal toxin reading frame dependent on a
 translational frameshift or read-through of a termination codon. With an appropriate
 low frequency of frame-shifting (eg 0.1 - 2%), it could even be sufficient to use a
 15 single transgene, if it were possible to synthesise the P7 portion and the toxin portion
 as overlapping genes. Upon assembly (which we have demonstrated in insect cells
 using the baculovirus vectors) and maturation, the protein precursors are cleaved and
 release the mature P7 and the toxin, which remain within the capsoids. These proteins
 are not released until capsoid disassembly occurs in insect gut cells. The processed
 20 form of the toxin is then able to kill the pest.

(c) HaSV particles devoid of nucleic acid carrying one or more suitable protein
 toxins and/or their mRNA

A protein toxin (or toxins) is expressed as a fusion with the capsid protein. The
 25 fusion protein then assembles into capsid carrying the toxin(s). These capsids present
 in the plant tissue exert an antifeeding effect on insects attaching the plant.

EXAMPLE 10

EXPRESSION OF HaSV IN OTHER DELIVERY VECTORS

30 Materials & Methods

(as indicated in earlier Examples)

Constructs similar to those for plant expression are introduced into yeast or bacteria by standard techniques. Virus particles are assembled for either fully infectious virus or any of the modified or biologically contained forms described in Example 9. Microbes produced in suitable fermentation or culture facilities and carrying such forms of the virus are then delivered to the crop by spraying. The microbial cell wall provides extra protection for the virus particles produced within the microbe.

Well established techniques exist for culture and transformation of yeast (Ausubel, F.M. *et al.* (eds) Current Protocols in Molecular Biology. J. Wiley & Sons, NY, 1989). An example of a yeast expression vector is pBM272, which contains the URA3 selectable marker (Johnston, M. & Davies, R.W. Mol. Cell. Biol. 4, 1440-8, (1984); Stone, D. & Craig, E. Mol. Cell. Biol. 10, 1622-32 (1990). Another example of an expression vector is pRJ28, carrying the Trp1 and Leu2 selectable markers. Yeast has recently been shown to support replication of RNA replicons derived from a plant RNA virus, brome mosaic virus (Janda, M. & Ahlquist, P. Cell 72, 961-70 (1993). Since the BMV replicase is distantly related to that of HaSV, and the two viruses are likely to replicate by similar strategies within cells, yeast cells probably contain all the cellular factors required for HaSV to generate infectious virus. For bacteria, suitable expression vectors have been described above.

EXAMPLE 11

The transvirus approach for insect pest control: Making transgenic plants expressing HaSV

1. Vector construction

A special binary vector was constructed for transforming plants with the HaSV genome. This vector is based on pART27 (A. Gleave (1992) Plant Mol.Biol.20, 1203-1207), which was modified to (1) carry an alternative origin of replication for the host

Agrobacterium tumefaciens and (2) incorporate restriction sites in the multiple cloning site for restriction enzymes Asc I and Pac I which recognise rare (8bp) sequences.

For engineering the multiple cloning site, pART27 was cut with SpeI and NotI. Ten
 5 picomoles of each of the two oligos whose sequence follows (TOP and BOTTOM)
 were annealed in 10 microlitres of water (heated to 80°C for 2 min and allowed to cool
 slowly to room temperature). The sticky ends on these annealed oligonucleotides
 allowed the insert to be cloned into pART27 (giving pART27mod) as described in
 Example No. 3 and 9.

10 Sequence of oligonucleotide:

TOP: 5'-GGCCGCTTAATTAAGGATCCGGCGCGCCA-3'

BOTTOM: 3'-CGAATTAATTCCTAGGCCGCGCGGTGATC-5'

(The PacI recognition sequence is TTAATTAA and that for AscI is GGCGCGCC). A
 15 4kbp SalI fragment from plasmid pART27mod (containing the right border, lacZ
 marker (+multiple cloning site)nptII gene for kanamycin resistance under control of
 the nos promoter and polyadenylation signal and the left border) was cloned into the
 13kbp vector pKT231 linearised with XhoI. Plasmid pKT231 carries the IncQ origin
 of replication for the host *Agrobacterium tumefaciens* and a resistance (marker) gene
 20 for streptomycin/spectinomycin. (Bagdasarian, M. & Timmis, K.N. (1982) Curr.
 Topics Microbiol. Immunol. **96**, 46-67). These two fragments were ligated using
 standard protocols (eg in Example No 3) and transformed into *E.coli* strain DH5α
 using standard protocols (eg in Example No 3). The resultant plasmid was named
 pJDML1.

2. Cloning HaSV genes into transfer plasmid

Construction of transfer vectors with HaSV genes:

- 5 Before the HaSV gene cassettes could be cloned into binary transfer vectors pART27 mod or pJDML1, they were re-cloned into the vector plasmid pBJ33 to provide flanking *AscI* and *PacI* sites. Plasmid pBJ33 (provided by Bart Janssen) is based on pBC SK(+) supplied by Stratagene), but with a multiple cloning site modified to contain the following sites:
- 10 *SacI/PacI/AscI/SacII/XbaI/SpeI/BamHI/PstI/EcoRI/EcoRV/HindIII/ClaI/SaII/XhoI/ApaI/PacI/AscI/KpnI*.

- The cDNA fragment corresponding to complete HaSV RNA 1 behind the 35S promoter and terminating in the hairpin cassette ribozyme and the CaMV
- 15 polyadenylation signal fragment (approx 6 kpb in total) was excised from plasmid pDHStuR1HC (Example 9) with *EcoRI* and cloned into *EcoRI*-cut vector pBJ33 to give plasmid pBJ33R1HC. Similarly, the cDNA fragment corresponding to complete HaSV RNA 2 behind the 35S promoter and terminating in the hairpin cassette
- 20 ribozyme and the CaMV polyadenylation signal fragment (approx 3.3 kbp in total) was excised from plasmid pDHStuR2HC (Example 9) as two fragments, one (covering the 35S promoter and the first 500 bp of the RNA 2 sequence) of about 1kbp with *EcoRI* and *R5rII* and the second (covering the remainder of the RNA 2 sequence, the ribozyme and the polyadenylation signal) of about 2.3kbp with *RerII* and *HindIII*.

These two fragments were simultaneously ligated into EcoRI and HindIII-cut vector pBJ33 to give plasmid pBJ33R2HC.

A 1.9 kbp fragment comprising the 5' 1.7 kbp of the HaSV capsid gene, together with the polyadenylation fragment, were excised from expression plasmid pDHVCAPB (described in Example 5) as a Eco RI - KpnI fragment and cloned into pTZ19U (pharmacia) cut with EcoRI and KpnI, giving pTZ19UEVCAPB., This portion of the HaSV capsid gene expression cassette was then re-excised as a HindIII-EcoRI fragment and cloned into PBJ33 cut with these enzymes. This plasmid (pBJ33EVCAPB) was then linearized with EcoRI and the ca. 800 bp EcoRI fragment from pDHVCAPB carrying the 35S promoter and the 5' 250 bp of the capsid gene inserted, followed by screening for orientation. The resulting plasmid carrying the reassembled complete capsid gene expression cassette was named pBJ33VCAPB.

Assembling binary plasmids:

The RNA 1 expression cassette was excised from plasmid pBJ33R1HC with AscI and PacI and cloned into pART27 mod cut with AscI and PacI to give pMLR1. The RNA 2 expression cassette was also cloned as an AscI-PacI fragment into pJDML1 cut with AscI and PacI to give pJDMLR2.

The capsid protein gene cassette was excised from pBJ33 VCAPB with PacI and cloned into plasmid pMLR1 cut with PacI. Resulting plasmids were screened for orientation and the plasmid with the capsid gene and RNA1 in the same orientation was named pMLR1V. The complete fragment carrying the HaSV capsid gene and

RNA 1 expression cassettes in pMLR1V was excised with AscI and cloned into pJDMLR2 linearised with AscI to give pHaSV1 (29kpb). This plasmid carries the HaSV capsid gene expression cassette and the HaSV RNA 1 and RNA 2 expression cassettes in this order and all in the same orientation. The kanamycin resistance gene is located upstream of the capsid gene and in the opposite orientation.

Table of constructs generated:

	Vector	Insert(s)	Name	#Plants (independent transformants)	Comments
10	pART27mod	RNA 1	pMLR1	15	control
	pJDML1	R1 + R2 +CAP	pHaSV1	30	complete virus
	pART27mod	R1 + CAP	pMLR1V	15	subvirus
	pJDML1	R1 + CAP	pJDMLR1V	30	subvirus
	pART27mod	RNA2	pMLR2	15	control
15	pJDML1	RNA2	pJDMLR2	15	control
	pART27mod	CAP	pMLVF	15	control

(CAP = HaSV capsid gene)

20 **3. Plant transformation and regeneration**

Binary transfer vectors (above) were transformed into *Agrobacterium tumefaciens* strain LBA4404 by electroporation (Lin, J.J. (1994) FOCUS **16**,18-19; Lin, J.J. (1994)

5

10

15

20

The leaf material was ground under liquid nitrogen to a powder and then extracted by further grinding in 0.45 ml NTES buffer (0.1 M NaCl, 10mM Tris-HCl pH 8.0, 1mM EDTA, 0.1% SDS) plus 0.45 ml Tris pH8.0-saturated phenol/chloroform. The slurry was vortexed, centrifuged for 3 min and the aqueous phase mixed with 1 volume of isopropanol to precipitate RNA and DNA. After resuspending the pellet in 0.1 ml

water, 1 volume of 4M LiCl was added and the mix stood on ice overnight before centrifugation to pellet RNA. The RNA was then analysed by gel electrophoresis according to the methods in Examples 1 and 2. HaSV specific RNAs were detected by Northern blottings as described in Example 2 and by using riboprobes made to detect
 5 the 3'-terminal 1000 nucleotides of each of RNA 1 and 2, made using the Promega Riboprobe kit and used as specified by the supplier.

For Southern blots: to detect HaSV genes in plant genomic DNA.

10 To recover plant genomic DNA, the supernatant from LiCl precipitation (above) was mixed with 2 volumes of ethanol. The pellet was redissolved and the DNA cut with BamHI before analysis on agarose gels and transfer to nylon membrane as described by Sambrook *et al* (1989) and by the manufacturer (Zetaprobe/BioRad). HaSV-specific bands were detected described above.

15

5. Bioassays on leaf material

Two small leaves (2-3 cm in length) were selected from each transformed plant selected, and placed in petri dishes containing 1.5% agarose in water. Three to 8
 20 neonate larvae were placed in each petri dish and observed for 3 days. At the end of this time, larvae were weighed and then total RNA extracted as described in Example 1. The extent of leaf damage was quantified by measuring the area of leaf consumed by each group of larvae over the three days of the assay (see Table 3).

Table 3**Preliminary bioassay of HaSV transgenic plants**

Three to 8 larvae were placed on a small leaf (from a newly regenerated plant) in a petri dish with no provision of fresh food, after 3 days, larvae were sacrificed and northern blotted; also, protein extracted from leaves of the plants were western blotted using anti-HaSV antisera.

Plant	Transform- ation Plasmid	Western Blot for HaSV capsid protein in plant (+/-)	Northern blot for HaSV RNA in plant	Larval Weight (mg)	Leaf Damage (mm ² consumed /larvae)
Negative Controls		-	-	2.1- 2.7±0.8*	61
1.1 (subvirus)	pJDMLR1V	+		1.1±0.2	29
(RNA1=p 71)					
3.2 (whole virus)	pHaSV1	+		1.0±0.4	38
3.4 (whole virus)	pHaSV1	+		1.2±0.4	32

* Diet was limiting (ran out of food) in some cases

Table 4**Further bioassay of HaSV transgenic plants**

Four - 6 individual larvae were fed leaf disc (50mm²) from control or transgenic plants at one disc each per day for 4 days, before transferred to artificial diet for a further 3 days. RNA was then extracted from the larvae and Northern blotting with HaSV-specific probes used to verify the presence of HaSV in the larvae.

15 Efficacy of HaSV as atransvirus in plants

Factors affecting the efficacy of HaSV are the viral dose required, the expression levels achieved in plants and the leaf damage observed. These need to be considered separately at this stage due to uncertainty about the efficiency of HaSV assembly in plants and because larvae will continue feeding for about one day after receiving a toxic dose of HaSV.

I. *Dose of Virus*

Infection with HaSV requires neonate larvae to eat up to 10 000 particles. Assuming that transgenic plants make only 1 particle per cell, this means the larvae must consume up to 10 000 leaf cells.

Since a small tobacco leaf contains about one million cells, larvae could acquire a toxic dose by consuming just 1% of the leaf.

This dose would correspond to as little as 0.000 000 5% of the soluble protein in these cells (330×10^{-13} g of HaSV per leaf in 7×10^{-13} g soluble plant protein per leaf).

II. *Expression levels*

Assuming standard levels of 1% expression and complete incorporation into virus particles, there should be about 10^8 particles per cell (7×10^{-9} g of protein per cell over 330×10^{-19} g per HaSV particle).

However, at present only part of this protein is likely to form infectious virus. If 1% does, then there would be 10^6 particles per cell, well above the toxic dose.

Initial results from Western blots suggest current expression at least exceed 0.1% of soluble cell protein. Processing of the precursor protein appears to occur to a variable extent, suggesting that particle assembly has also occurred.

The dose of infectious virus delivered by transgenic plants must be quantified by appropriately standardised bioassays.

Optimisation of the infectious virus level will be achieved by improving virus assembly rather than just boosting expression of components - this represents a fundamental difference to the situation with toxins like Bt.

I. Dose of Virus

Infection with HaSV requires neonate larvae to eat up to 10 000 particles. Assuming that transgenic plants make only 1 particle per cell, this means the larvae must consume up to 10 000 leaf cells.

Since a small tobacco leaf contains about one million cells, larvae could acquire a toxic dose by consuming just 1% of the leaf.

This dose would correspond to as little as 0.000 000 5% of the soluble protein in these cells (330×10^{-13} g of HaSV per leaf in 7×10^{-13} g soluble plant protein per leaf).

II. Expression levels

Assuming standard levels of 1% expression and complete incorporation into virus particles, there should be about 10^8 particles per cell (7×10^{-9} g of protein per cell over 330×10^{-19} g per HaSV particle).

However, at present only part of this protein is likely to form infectious virus. If 1% does, then there would be 10^6 particles per cell, well above the toxic dose.

Initial results from Western blots suggest current expression at least exceed 0.1% of soluble cell protein. Processing of the precursor protein appears to occur to a variable extent, suggesting that particle assembly has also occurred.

The dose of infectious virus delivered by transgenic plants must be quantified by appropriately standardised bioassays.

Optimisation of the infectious virus level will be achieved by improving virus assembly rather than just boosting expression of components - this represents a fundamental difference to the situation with toxins like Bt.

III. *Leaf damage*

While as little as 1% of the leaf (and more likely far less) may be sufficient to deliver a toxic dose of HaSV, larvae will keep feeding for a limited period after becoming infected. This makes it necessary to determine the extent of leaf damage empirically.

Our initial observations were that plants making detectable levels of HaSV capsids showed reduced susceptibility to larval feeding; this has not been quantified yet, and the assay was a severe one.

Consumption of leaf material by infected larvae may be estimated indirectly using our data on larval growth and frass production, which are approximately equal. Since neonate frass production is too low to quantify, the data were obtained from 4-day old larvae. These produce 30 mg of frass over 7 days, compared to 400 mg for uninfected controls. Neonate growth and frass production may be estimated at 10% of this figure.

Assuming that 1 mg growth or frass = 3 mg leaf material, an infected neonate will consume about 5% of a small tobacco leaf (20 mg of a total fresh weight of 350 mg) over seven days compared to over 60% for an uninfected control (240 mg of 350 mg).

Biosafety Considerations

It is believed that the approach of controlling pests by making an insect virus in transgenic plants is not dangerous to the environment. This is despite our very tentative observation that some HaSV replication is observed in protoplasts. There has been widespread debate recently concerning the safety of protecting crops against plant viruses by inserting transgenes expressing viral proteins into the plants. Falk, B.W. and Bruening, G., 1994 (*Science* **263**, 1395-1396) identified 3 possible mechanisms which might result in the appearance of novel viruses. These mechanisms are transencapsidation, phenotypic mixing and heterologous recombination.

Plasmids	Bioassays weights (mg) of larvae fed on protoplast extracts	HaSV RNAs 1 & 2 detected by Northern blotting of RNA extracted from larvae
1. pDHStuR1HC + pDHVCAPB	29 ± 15	+
2. pDHStuR1 HDV + pHV CAPB	57 ± 25	(-)
3. Control: (diet only/diet + protoplasts)	85 ± 15	-
4. pDHStuR1 HC + pDHStuR2 HC + pDHVCAPB	33 ± 28	+
5. pDHStuR1 HDV + pDHStuR2 HDV + pDHVCAPB	64 ± 22	-

- ii) RNA extraction from larvae showed
- that larvae fed protoplasts transfected with pDHStuR1HC + pDHStuR2HC + DHVCAPB contained both RNA1 and 2 of HaSV in intact form.
 - that larvae fed protoplasts transfected with pDHStuR1HC + DHV CAPB (subvirus) contained a very small amount of intact HaSV RNA1 and a considerably greater amount of degraded RNA1.
 - that larvae fed protoplasts transfected with pDHStuR1HDV + pDHStuR2HDV + pDHVCAPB contained no HaSV RNA with one exception.
 - that larvae fed protoplasts transfected with pDHStuR1 HDV + pDHVCAPB contained no HaSV RNA.

Conclusions:

The HC (HaSV expression) constructs with the hairpin cassette ribozyme give infectious particles with both RNAs; the HDV expression constructs do not under these conditions.

That the subvirus approach results in RNA1 replicating in larvae but this RNA is degraded because it cannot be encapsidated due to missing replicatable RNA2.

- That subvirus approach gives stunting as effectively as does the complete virus approach under these conditions.

[illegible]

EXAMPLE 12**CAPSOVECTOR**

5 The aim of this section is to describe the invention of capsovectors and present supporting data. In addition ideas for their further improvement will be presented.

10 The capsovector is a virus-like particle (VLP) from a small RNA insect virus with a pest insect host whose properties will facilitate the entry (or vector) into the cells of the pest insect either, or both, RNA or a protein that will induce toxicity in the insect. Capsovectors can be produced in transgenic crop plants targeted by the host insect pests or produced for spray applications by transgenic plants or recombinant microorganisms.

15 There are two types of capsovectors, ones that vector RNA moieties (RNA capsovectors) and ones that vector proteinaceous moieties (protein capsovectors). Because of their distinct properties, RNA and protein capsovectors will be dealt with separately. However, the description of the protein capsovector will rely heavily on the preceding description of the RNA capsovector.

20 In this invention, the *Helicoverpa armigera* stunt virus (HaSV) of the Tetraviridae will be used as the model insect small RNA virus. However, this does not exclude other types of insect small RNA viruses being used in a similar manner.

25 **Characteristics of HaSV pertinent to capsovectors.**

HaSV is a member of the Tetraviridae family of viruses and infects only midgut cells of young larvae of heliothine insects (Hanzlik et al., 1993) after ingestion with food. Numerous attempts to grow the virus in non-gut tissue, other insects and cultured cells
30 have failed. This believed to be due to the ability of the HaSV capsid protein to bind and enter only midgut cells of the host insect (Hanzlik et al., 1995).

HaSV has been characterised in great detail at the molecular level. Its physical characteristics have been determined (Hanzlik et al., 1993) and its complete genome
35 sequenced (Gordon, et al., 1995; Hanzlik, et al., 1995).

particles is aimed at replication of new virions replication with any toxicity is a by-product of this process; that contained within capsosvectors is aimed at rapid toxicity to the cell and not at replicating new capsosvectors.

- 5 Important to the toxic activity of a RNA capsosvector are the properties of viral genes that lead to the activities of encapsidating, protecting, entry, uncoating and translating of the viral RNA in the host cell. These activities are necessary for successful virus infection yet do not involve replication. Because of the simplicity of HaSV, all of these properties are contained in one gene and its product, the coat protein, p71.
- 10 Generally, the coat protein gene of other small RNA viruses have the same capacity and therefore are also suitable for capsosvectors. The p71 gene of HaSV is employed in illustrating the capsosvector invention and its properties are explained and demonstrated under their respective headings below.

- 15 The capsosvector adds toxicity to the virus functionalities by using sequences exogenous to HaSV or viruses on the RNA contained within the capsosvectors. These toxic sequences are aimed at inducing rapid and direct toxicity to the cell the capsosvector has entered. The sequences are either translated into protein or fold the RNA into appropriate secondary structures which then causes a toxic lesion in the cell.
- 20 These sequences are explained below as well.

- It must be pointed out here that viruses avoid inducing rapid and direct toxicity to cells they have entered as this is deleterious to viral replication. Because the cell must be viable for the production of new viruses, many viruses do not induce cell death until
- 25 late in their infections if at all. Indeed, infection of hosts by many insect small RNA viruses different from HaSV does not result in any discernible response in the host insect upon its being infected. These innocuous viruses can still be exploited, however, because the viruses are able to perform the previously mentioned functions of encapsidation and protection of the labile RNA genome, ability of the particles to
- 30 enter host cells, uncoating and subsequent translation of the RNA genome. Appropriate placement of toxin sequences with those of the coat proteins of these viruses as described in this invention can result in toxicity and subsequent control of the insect pest host.

- 35 **Encapsidation.** Encapsidation is defined here as the process of forming a virus particle or a virus-like particle that incorporates RNA into its interior. For this process

to occur, interactions among the capsid proteins (encoded by the RNA) and between the proteins and specific regions of the RNA must occur that result in the ordered aggregation of 240 copies of the coat protein and the RNA strand(s). The specific regions on the RNA are either or both a defined primary sequence or a specific secondary structure arrived at by a number of primary sequences.

All three types of interactions, protein/protein, RNA/protein and RNA folding into secondary structures, reside in the capsid gene ORF of HaSV. This is shown by the following data:

1. When the ORF the HaSV capsid protein is expressed in insect cells with a recombinant baculovirus, VLPs can be observed with transmission electron microscopy (TEM) in sections of fixed, positively-stained, infected cells. They are highly similar in morphology and dimensions to native virions observed in gut tissue of diseased insects. The VLP morphology is that of a smooth surfaced sphere with a diameter of 35-40 nm identical to virus particles observed in diseased tissue. Also noted is that a fraction of the particles have dark, electron dense cores. The fraction of VLPs having electron dense cores is smaller than that observed for virus particles observed in diseased tissue. The electron dense cores indicate that the particles contain electron dense, non proteinaceous RNA.

Methods and Materials: A recombinant baculovirus able to express p71 was made by the following procedure. An amplicon containing the p71 ORF was obtained from a PCR reaction made with HP64NEUK (GGCCGGATCCAGACATGCTGGAGTGGCGTCAC) and HVP6C2 (GGGATCCCTAATTGGCACGAGCGGCGC) off a DNA template consisting of the pT7T2p71 plasmid used for the bacterial expression of the capsid gene. This fragment contains the p71 ORF flanked with BamHI sites and the initiating AUG placed behind a more favourable context for expression in eukaryotes. This fragment was restricted with BamHI and cloned into the baculovirus expression vector, pVL941, which was transfected into Sf9 cells with linearized wildtype baculovirus DNA. A recombinant baculovirus, Bacp71, was obtained by standard means (King and Possee, 1992).

For TEM, Sf9 cells infected for three days with Bacp71, were harvested, fixed with gluteraldehyde, embedded in LRWhite resin and sectioned for

examination with TEM using standard procedures. The sections were examined with a JEOL 100CX transmission electron microscope.

2. When the VLPs are purified and characterised, they also show highly similar characteristics to native HaSV virions. Buoyant densities of the particles are similar 1.296 g/ml for virions and 1.29 g/ml for VLPs. Particle morphologies as seen by negative stained TEM are highly similar with the micrographs for both virions and VLPs showing spheres of 35-40 nm spheres possessing electron dense interiors. The electron dense interiors also indicate the presence of RNA.

Methods and materials. After a three day infection with Bacp71 in a 220 cm² flask, Sf9 cells were harvested, pelleted, washed, and lysed with a freeze-thaw cycle in 10 ml of buffer. This was centrifuged at 10,000 x g for 10 min and the supernatant was recovered and recentrifuged on top of a 30% sucrose cushion. The pellet was then redissolved in 0.5 ml of buffer (50mM Tris pH7.5, 5 mM CaCl) overnight and placed on top of a solution in a SW41 tube (Beckman) consisting of 5 mls each of 60% and 30% of CsCl in buffer. This was centrifuged overnight at 40,000 rpm and the sole band located in the middle of the tube was removed, placed into 10 ml of buffer and pelleted by centrifugation at 35,000 rpm in a SW41 rotor. The pellet was dissolved in 50 ml of buffer.

VLPs in the solution were examined on a JEOL 100CX microscope with standard procedures after negative staining with uranyl acetate..

3. The VLPs contained RNA. The above observations indicated that the VLPs in the were electron dense and that RNA was within the VLPs. These observations were confirmed when the VLPs were extracted for RNA and the RNA analysed by agarose gel electrophoresis. The RNA was 2900 bases in length, the expected size of the mRNA transcribed from the baculoviral genome. When the RNA was probed with a radioactively labelled probe specific for p71 sequences, strong hybridisation occurred, showing that the RNA was the p71 mRNA.

Methods and materials. RNA was removed from the purified VLPs with extraction with phenol, phenol/CH₃Cl and CH₃Cl and ethanol precipitation. This was then run on an 1% formaldehyde agarose gel. A northern blot of the gel was done with standard procedures. The ³²P-

labelled probe was prepared from clone HR326 which contains the p71 sequence.

4. The p71 mRNA extracted from the VLPs are shown to have heterologous, non-HaSV sequences. This is shown by the recovered RNA strand having a length of 2900 bases which is equal to the predicted size of the mRNA transcribed from the baculoviral genome and expressing the p71 ORF. Only 1900 bases of the mRNA are sequences from the p71 ORF, the rest being sequences from the baculovirus expressing the gene. When a radioactively labelled probe specific for the polyhedrin gene located between the p71 gene the polyadenylation signal for the polyhedrin gene was hybridized to the RNA extracted from the VLPs, a strong hybridization signal was seen on the autorad. This shows that the signals required for encapsidating RNA are present in the p71 ORF and that non-HaSV sequences can be placed inside the VLP.
Methods and materials. RNA extracted from VLPs were northern blotted by standard procedures and probed with a labelled 950 bp Hind III fragment from the baculoviral transfer vector pVL941 having sequences 3' to the inserted p71 gene and 5' to the polyadenylation site of the polyhedrin gene.
5. The RNA having the capsid protein ORF is specifically encapsidated. No other RNA except for the p71 mRNA is present in the VLPs. This is shown by the failure of another highly transcribed region of the baculoviral genome (therefore also present in great abundance inside the baculoviral infected cell) failing to hybridise to the VLP RNA. When a probe specific for the p10 mRNA, a late gene product from the baculovirus, is hybridised to the RNA extracted from the purified VLPs, no signal occurs on the northern blot.
Methods and materials. RNA extracted from VLPs were northern blotted by standard procedures and probed with a ³²P-labelled 245 bp Xba I/Eco RI fragment from the baculoviral transfer vector pAcUW31 having sequences 3' to the start of transcription of the p10 promoter for AcMNPV baculovirus.

Thus RNA sequences contained within the ORF of HaSV capsid protein (p71) can produce VLPs that encapsidate only RNAs having the p71 sequence. If there are exogenous sequences not from HaSV also on the p71 mRNA these then can be

encapsidated. This shows that sequences for toxins can be placed within the VLP if it possesses a p71 sequence.

Protection. Normally labile RNA contained within the capsowector must be protected from degradation during the period between its encapsidation inside the cell where it was produced and its entry into the cell where it will effect toxicity. This is particularly important because part of this period is spent in the insect gut that is a highly degradative milieu for both RNA and protein. This function is performed by the 240 copies of the capsid protein which form a protective shell around the virion RNA.

The protective properties of the HaSV coat protein for both the virion and the VLP was shown by western blots of the capsid protein from HaSV particles, HaSV VLPs, and lipophorin before and after timed exposure to the gut contents of heliothis larvae: The data shows that:

1. When a non-viral globular protein like lipophorin is exposed to the contents of the heliothis midgut, rapid degradation of the protein occurs within 10 minutes.
 2. When either an HaSV virion or VLP is exposed in the same manner, minimal degradation of the protein occurs despite extended exposure times (>2 hr).
- Thus, when translated, the RNA sequences contained within the p71 ORF lead to protection of the protein and RNA of the VLP.

Methods and materials. The midguts of fifth instar *Heliothis armigera* were excised, their contents removed and centrifuged at 14,500 x g for 15 min. Into 10 ml of the contents were placed 1 ml of a solution having 1 mg protein of either lipophorin, HaSV virions or HaSV VLPs. At timed intervals 1 ml of this solution was removed and immediately boiled in SDS-PAGE sample buffer. SDS-PAGE and immunoblots with the respective antisera were then carried out with standard procedures.

Cell Entry. Studies of animal host/virus systems have shown that entry of virions into cells is mediated by cellular receptors located on their exteriors and viral acceptor proteins or VAPs (Lentz, 1990). The HaSV VAP is by exclusion of any other possibility, p71, it being the sole protein component of the HaSV capsid (Hanzlik et al., 1993, 1995). In particular, it is believed that a specific region of p71, between residue 274 and 439 of the protein sequence, is responsible for the binding of HaSV to the presently unknown host cell receptor.

5

10

15

20

25

30

35

produced from an RNA having a translatable p71 encapsidates an RNA with a nontranslatable p71 sequence but possessing the encapsidation signal that ensures the encapsidation of the strand with a translatable toxin sequence.

2. IRESs. IRESs are one of the exceptions to the general rule noted above for translation. They are sequences on an RNA strand that allow ribosomes to bind an RNA internally instead of at the 5' cap to initiate protein translation. They are located in the 5' regions of picornavirus vRNAs and specialised genes from certain organisms (Jackson et al., 1994)). They have been employed to allow the translation of two proteins from the same mRNA (Lipsick and Smarda, 1993) and thus can be employed to express a toxin from a sequence located. after a translatable p71. An advantage to this approach is that many IRESs are host specific. If the IRES used in the RNA capsovector is able to function only in the target organism, the toxin is produced only in the target organism and not in the organism producing the RNA capsovector. A picornavirus with a targeted pest insect host will possess a suitable IRES for a capsovector. For example, an IRES from a picornavirus with a *Heliothis* host will be used in a capsovector constructed with p71 from HaSV and a cytotoxin such as the ricin A fragment. This particular capsovector will affect only heliothine insects and no others as well as not producing a ricin A fragment in the plant producing the capsovector.

Translation into Toxicity.

Ultimately, all of the above abilities of encapsidation, protection, entry, and uncoating must be induced or accomplished by sequences in the RNA strand that produce the capsovector. Also important are sequences on the RNA, derived from other viral and non-viral sources as well as from HaSV, that are responsible for the toxic activity and if required, translation into protein that confer the toxicity. These types of sequences will be dealt with in separate sections.

RNA sequences leading to toxicity of the organism or cell can either be translated into protein or the sequences cause secondary structures to be made on the RNA strand that lead to toxicity.

1. Toxicity from protein sequences. Both types of toxins, nerve toxins, specific for insects and work at the level of the organism (Binnington and Baule, 1993) and cytotoxins (Stripe and Barbeiri, 1986), toxic only to the cell the capsovector has entered, can be used with RNA capsovectors. However, the former type of toxin will

have to have a secretion signal appropriate to midgut cells. Examples of these proteins are described in Maeda et al., 1991.

2. Toxicity from RNA secondary structure sequences. RNA secondary structures will have the ability to cause toxicity to the cell to which they have been vectored by RNA capsovectors. These structures are caused by primary sequences; however, any number of primary sequences on the RNA strand can lead to the same toxic secondary structure. There are three types of these sequence structures that will be appropriate for RNA capsovectors: antisense sequences, ribozymes and mimicking structures. The first two types are reviewed by Eguchi et al., 1991 and elsewhere herein, respectively and aimed at preventing the expression of key cellular enzymes. The latter is novel and will be detailed.

- It is the activity of the HaSV replicase and not virosis or accumulation of viruses that causes the midgut cell to cease functioning. This is shown by data generated from the following experiment. When protoplasts are transfected with genes that make a replicatable RNA1 and only the capsid protein and not a replicatable RNA2 (R1-HC and VCAPB according to procedures listed above), stunting occurs. When the stunted larvae are extracted for RNA which is then northern blotted with probe for HaSV nucleic acid, only RNA1 of HaSV is seen to be present. Stunting does not occur when the protoplasts are transfected with genes that do not make a replicatable RNA1 (lacking an effective ribozyme to cleave after the last viral base in the gene) and only the capsid protein and not a replicatable RNA2 (R1-HDV and VCAPB according to procedures listed elsewhere in patent). When the stunted larvae are extracted for RNA which is then northern blotted with probe for HaSV nucleic acid, no HaSV RNA is seen to be present.
- These data are consistent with RNA1 being encapsidated and able to enter midgut cells of the larvae. RNA1 is able to self-replicate but not produce virions as there is no replicatable RNA2 which has the p71 ORF. The self-replication leads to antibiosis due to apoptosis of the midgut cells having the replicating RNA1. The particles made in the same manner but having RNA1 not able to self-replicate (due to the 3' sequences left by the defective ribozyme) were unable to stunt the larvae.
- As shown in other systems, the activity of the replicase may be mimicked by placing a by-product of replicase activity into the cell. This action "tricks" the cell into initiating anti-viral measures such as shut-down of protein synthesis or cell death by apoptosis (the activity believed to be responsible for the antibiosis caused by HaSV).

- 10

15

20

In concept, protein capsosvectors are similar to the immunotoxins successfully used in human cancer therapy where cytotoxic moieties are vectored to specific cells by fusion to a specific binding moiety such as antibodies or cytokines. By themselves, the cytotoxin fragments do not display toxicity even when injected intravenously. Only when they are attached or fused to a binding element does cytotoxicity occur to those cells to which the element binds. The binding element of the protein capsosvector is the VAP part (see above) of the HaSV VLP which is able to bind and enter midgut cells of heliothis larvae.

35

cell. The toxin fragment will be contained within the capsid shell until the capsovector enters the midgut cell. Also, it is the capacity of capsovectors to protect the toxin moiety from degradation in the midgut lumen that makes it distinct from other insect control factors that are fused to elements that only "interact" with midgut cells.

5

In the following sections, each of the two elements of a capsovector, the cytotoxic fragment and a viral coat protein gene, will be described followed by descriptions of how the capsovector are constructed with data supporting their feasibility.

- 10 **The toxin.** Several cytotoxin fragments suitable for protein capsovectors are available in readily accessible form. , They have extensive literature describing their activity in various fusions for use in immunotoxins (Thorpe et al., 1982). For expression in plants, plant-derived fragments of proteinaceous toxins, such as ricin A fragment, which are not toxic to plant cells but toxic to animal cells will be the most suitable.
- 15 For expression in microorganisms, toxins of bacterial origin may be the most suitable, it being that they are not toxic to the microorganism producing the protein capsovectors.

- 20 As described in the section on RNA capsovectors, p71 has the ability to self-assemble into VLPs when expressed in various non-host expression systems. Also shown was the ability of the VLPs and virions to resist degradation. and bind to, then enter, midgut cells. The critical question concerning the feasibility of the protein capsovector using p71, the HaSV coat protein, is the ability of VLPs to form with the modified coat proteins fused to toxins.

25

Construction of model protein capsovectors with capsid protein, p71 and a reporter peptide.

- The cytotoxin fragment can be fused to p71 in a number of ways. In order to test the possibilities, a reporter peptide fragment was used in place of the toxins. This allows a more rapid characterisation of the products as immunodetection
- 30 of the exogenous fragment is facilitated by a commercially available monoclonal antibody specific for the fragment (IBI FLAG Biosystem). Two constructs were made using standard techniques. The FlagT construct placed the reporter fragment, sized 2501 Da, at the C-terminus of p71 (Fig. 8). The FlagM construct placed a reporter fragment, sized 1243 Da, in the middle of the p71 sequence near the site where p71 is
- 35 cleaved into p64 and p7 (Fig. 8). Both constructs should have resulted in the reporter fragment being placed into the interior of the VLP where p7 is located.

Formation of VLPs made with modified p71. Recombinant baculoviruses were made with the modified p71 genes by standard techniques (King and Possee, 1992) and used to infect Sf9 cells. When these cells were processed and examined as before with TEM, particles highly similar to VLPs made with unmodified p71 and HaSV virions were observed. The particles were purified on CsCl gradients and were examined with SDS-PAGE, immunoblots, negatively stained TEM. In addition, the particles were tested for their ability to protect the FLAG epitope on the fused peptide fragment. These experiments showed:

1. The purified particles were highly similar to HaSV virions and VLPs made with unmodified p71 and showed the characteristic 35-40 nm diameter spheres with a fraction having the electron dense cores. The FlagT particle possessed a buoyant density of 1.31 g/ml compared to 1.29 g/ml for unmodified VLPs and virions.
2. Proteins extracted from the particles and observed with SDS-PAGE had the molecular weights predicted from the constructs (1243 Da and 2501 Da for FlagM and FlagT respectively. In addition, post-assembly cleavage into p64 was observed for the FlagM particle. No processing was seen in the FlagT particle.
3. Proteins extracted from the particles reacted with both the anti-p71 antisera and with the FLAG monoclonal antibody on immunoblots.
4. The FLAG epitope was protected during exposure to heliothis midgut contents. This is shown by the FLAG epitope remaining at the original molecular weight and therefore undegraded.

Hybrid VLPs. Although VLPs can be made with only modified p71 fused to the reporter peptide, and protection of the exogenous reporter peptide occurs, a protein capsovector made with both native p71 and modified p71 fused to a cytotoxin may function better. At present it is not clear what properties of the "native" VLP, if any, are altered with the addition of the exogenous, fused peptides to p71. If any deleterious properties arise such as poor stability of the particle, a resolution to the problem will be to produce a hybrid particle. This will minimise any disruption of desirable properties of the native VLP by the modified p71.

Hybrid VLP expression in plants. Three ways can be envisioned to express, either in a transgenic plant, the hybrid capsovector which requires two distinct, but closely

related proteins. The most obvious is two insert two genes into the plant, one for each protein. However, the use of elements from plant viruses can make it possible to express a capsovector from a single gene. These elements are suppressible stop contexts and frame-shift sequences that are detailed by Sleat and Wilson (1992). The use of these elements make it feasible to precisely regulate the ratio of coat protein to coat protein-toxin fusion.

References

- 10 Agrawal, D.K. & Johnson, J.E. (1992) Sequence and analysis of the capsid protein of *Nudaurelia capensis* ω virus, an insect virus with T=4 icosahedral symmetry. *Virology* 190, 806-804.
- Agrawal, D.K. & Johnson, J.E. (1995) Assembly of the T=4 *Nudaurelia capensis* ω virus capsid, post-translational cleavage, and specific encapsidation of its mRNA in a baculovirus expression system. *Virology* (in press)
- 15 Binnington, K.C. and Baule, V.J. (1993) Naturally occurring insecticidal molecules as candidates for genetic engineering. in *Molecular approaches to fundamental and applied entomology* (Oakeshott, J. and Whitten M., eds) Springer-Verlag, New York.
- Eguchi, Y., Itoh, T., and Tomizawa, J. (1991) Antisense RNA. *Annual Reviews of Biochemistry* 60, 631-652
- 20 Gallagher, T.M. and Rueckert, R.R. (1988) Assembly-dependent maturation cleavage in provirions of a small icosahedral insect ribovirus. *Journal of Virology*, 62, 3399-3406.
- Gordon, K.H.J., Johnson, K.N. & Hanzlik, T.N. The larger genomic RNA of
- 25 *Helicoverpa armigera* stunt virus encodes the viral RNA polymerase and has a novel 3'-terminal tRNA-like structure. *Virology* (in press)
- Hanzlik, T.N., Dorrian, S.J., Gordon, K.H.J., & Christian, P.D. (1993) A novel small RNA virus isolated from the cotton bollworm, *Helicoverpa armigera*. *Journal of General Virology* 74, 1805-1810
- 30 Hanzlik, T.N., Dorrian, S.J., Gordon, K.H.J., & Christian, P.D. Sequence of RNA2 of the *Helicoverpa armigera* stunt virus (Tetraviridae) and bacterial expression of its genes. *Journal of General Virology* (in press).
- Jackson, R.J., Hunt, S.L., Gibbs, C.L. and Kaminski, A. (1994) Internal initiation of translation of picornaviral RNAs. *Molecular Biology Reports* 19, 147-159
- 35 King, L.A. and Possee, R.D. (1992) *The Baculovirus Expression System*. Chapman and Hall, London

- 30

(1) GENERAL INFORMATION:

(ia) INVENTORS: P. D. CHRISTIAN, K. H. J. GORDON and
T. N. HANZLIK

```

15      (iii) NUMBER OF SEQUENCES: 53

```

20 (A) ADDRESSEE: DAVIES COLLISON CAVE
(B) STREET: 1 LITTLE COLLINS STREET
(C) CITY: MELBOURNE
(D) STATE: VICTORIA
(E) COUNTRY: AUSTRALIA
(F) ZIP: 3000

(A) MEDIUM TYPE: Floppy disk
(B) COMPUTER: IBM PC compatible
(C) OPERATING SYSTEM: PC-DOS/MS-DOS
(D) SOFTWARE: PatentIn Release #1.0, Version #1.25

(A) APPLICATION NUMBER:
(B) FILING DATE: 13 AUGUST 1993
(C) CLASSIFICATION:

35
(viii) ATTORNEY/AGENT INFORMATION:

5 (ix) TELECOMMUNICATION INFORMATION:
(A) TELEPHONE: (613) 254 2777

5

(A) LENGTH: 13 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

15

13

20

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 28 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

25

(ii) MOLECULE TYPE: DNA

30

(x1) SEQUENCE DESCRIPTION: SEQ ID NO:2:

28

35

(2) INFORMATION FOR SEQ ID NO:3:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 27 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(11) MOLECULE TYPE: DNA

(x1) SEQUENCE DESCRIPTION: SEQ ID NO:3:

ATGGAGGATG CTGGAGTGGC GTCACAG

27

(2) INFORMATION FOR SEQ ID NO:4:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 27 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(11) MOLECULE TYPE: DNA

(x1) SEQUENCE DESCRIPTION: SEQ ID NO:4:

ATGAGCGAGG CCGGCGTCGC GTCACAG

27

(2) INFORMATION FOR SEQ ID NO:5:

5

(B) TYPE: nucleic acid

(D) TOPOLOGY: linear

10

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:5:

30

(2) INFORMATION FOR SEQ ID NO:6:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 31 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:6:

CCATCGATGC CGGACTGGTA TCCCGAGGGA C

31

(2) INFORMATION FOR SEQ ID NO:7:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 39 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:7:

CCATCGATGA TCCAGCCTCC TCGCGGCGCC GGATGGGCA

39

5

GGAGATCTAC ATATGGGAGA TGCTGGAGTG

10

(i) SEQUENCE CHARACTERISTICS:

(B) TYPE: nucleic acid

15

(ii) MOLECULE TYPE: DNA

20

GTAGCGAACG TCGAGAA

17

25

(1) SEQUENCE CHARACTERISTICS:

30

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

35

(ii) MOLECULE TYPE: DNA

5

31

10

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:20:

28

25

(2) INFORMATION FOR SEQ ID NO:21:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 29 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:21:

AATTACATAT GCGGCCGCC GTTCTGCC

29

(2) INFORMATION FOR SEQ ID NO:22:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 29 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:22:

AATTACATAT GTTCGCGGCC GCCGTTTCT

29

(2) INFORMATION FOR SEQ ID NO:23:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 19 amino acids

(B) TYPE: amino acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein - N terminal

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:23:

Phe Ala Ala Ala Val Ser Ala Phe Ala Ala Asn Met Leu Ser Ser Val

1

5

10

15

Leu Lys Ser

(2) INFORMATION FOR SEQ ID NO:24:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 20 amino acids

(B) TYPE: amino acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein - internal

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:24:

Pro Thr Leu Val Asp Gln Gly Phe Trp Ile Gly Gly Gln Tyr Ala Leu

1

5

10

15

Thr Pro Thr Ser

Parameter	Value	Unit
Initial concentration	1.0	g/L
Initial pH	7.0	
Temperature	25	°C
Time	0, 1, 2, 4, 8, 16, 32, 64, 128, 256, 512, 1024, 2048, 4096, 8192, 16384, 32768, 65536, 131072, 262144, 524288, 1048576, 2097152, 4194304, 8388608, 16777216, 33554432, 67108864, 134217728, 268435456, 536870912, 1073741824, 2147483648, 4294967296, 8589934592, 17179869184, 34359738368, 68719476736, 137438953472, 274877906944, 549755813888, 1099511627776, 2199023255552, 4398046511104, 8796093022208, 17592186044416, 35184372088832, 70368744177664, 140737488355328, 281474976710656, 562949953421312, 1125899906842624, 2251799813685248, 4503599627370496, 9007199254740992, 18014398509481984, 36028797018963968, 72057594037927936, 144115188075855872, 288230376151711744, 576460752303423488, 1152921504606846976, 2305843009213693952, 4611686018427387904, 9223372036854775808, 18446744073709551616, 36893488147419103232, 73786976294838206464, 147573952589676412928, 295147905179352825856, 590295810358705651712, 1180591620717411303424, 2361183241434822606848, 4722366482869645213696, 9444732965739290427392, 18889465931478580854784, 37778931862957161709568, 75557863725914323419136, 151115727451828646838272, 302231454903657293676544, 604462909807314587353088, 1208925819614629174706176, 2417851639229258349412352, 4835703278458516698824704, 9671406556917033397649408, 19342813113834066795298816, 38685626227668133590597632, 77371252455336267181195264, 154742504910672534362390528, 309485009821345068724781056, 618970019642690137449562112, 1237940039285380274899124224, 2475880078570760549798248448, 4951760157141521099596496896, 9903520314283042199192993792, 19807040628566084398385987584, 39614081257132168796771975168, 79228162514264337593543950336, 158456325028528675187087900672, 316912650057057350374175801344, 633825300114114700748351602688, 1267650600228229401496703205376, 2535301200456458802993406410752, 5070602400912917605986812821504, 10141204801825835211973625643008, 20282409603651670423947251286016, 40564819207303340847894502572032, 81129638414606681695789005144064, 162259276829213363391578010288128, 324518553658426726783156020576256, 649037107316853453566312041152512, 1298074214633706907132624082305024, 2596148429267413814265248164610048, 5192296858534827628530496329220096, 10384593717069655257060992658440192, 20769187434139310514121985316880384, 41538374868278621028243970633760768, 83076749736557242056487941267521536, 166153499473114484112975882535043072, 332306998946228968225951765070086144, 664613997892457936451903530140172288, 1329227995784915872903807060280344576, 2658455991569831745807614120560689152, 5316911983139663491615228241121378304, 10633823966279326983230456482242756608, 21267647932558653966460912964485513216, 42535295865117307932921825928971026432, 85070591730234615865843651857942052864, 170141183460469231731687303715884105728, 340282366920938463463374607431768211456, 680564733841876926926749214863536422912, 1361129467683753853853498429727072845824, 2722258935367507707706996859454145691648, 5444517870735015415413993718908291383296, 10889035741470030830827987437816582766592, 21778071482940061661655974875633165533184, 43556142965880123323311949751266331066368, 87112285931760246646623899502532662132736, 174224571863520493293247799005065324265472, 348449143727040986586495598010130648530944, 696898287454081973172991196020261297061888, 1393796574908163946345982392040522594123776, 2787593149816327892691964784081045188247552, 5575186299632655785383929568162090376495104, 11150372599265311570767859136324180752990208, 22300745198530623141535718272648361505980416, 44601490397061246283071436545296723011960832, 89202980794122492566142873090593446023921664, 178405961588244985132285746181186892047843328, 356811923176489970264571492362373784095686656, 713623846352979940529142984724747568191373312, 1427247692705959881058285969449495136382746624, 2854495385411919762116571938898990272765493248, 5708990770823839524233143877797980545530986496, 11417981541647679048466287755595961091061972992, 2283596308329	

(2) INFORMATION FOR SEQ ID NO:25:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 6 amino acids
 (B) TYPE: amino acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein - internal

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:25:

Phe Ala Ala Ala Val Ser

1 5

(2) INFORMATION FOR SEQ ID NO:26:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 23 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: RNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:26:

GCGCCCCUG GGAUACCAGG AUC

23

(2) INFORMATION FOR SEQ ID NO:27:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 17 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(x1) SEQUENCE DESCRIPTION: SEQ ID NO:27:

TCAGCAGGTG GCATAGG

17

(2) INFORMATION FOR SEQ ID NO:28:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 32 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(ix) FEATURE:

- (A) NAME/KEY: CDS
(B) LOCATION: 6..32

(x1) SEQUENCE DESCRIPTION: SEQ ID NO:28:

CCCAT ATG GGC GAT GCC GGC GTC GCG TCA CAG

32

Met Gly Asp Ala Gly Val Ala Ser Gln

1

5

5

(A) LENGTH: 9 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

10

(x1) SEQUENCE DESCRIPTION: SEQ ID NO:29:

15

1 5

20

(A) LENGTH: 32 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

25

(A) NAME/KEY: CDS

(B) LOCATION: 6..32

35

CCCAT ATG AGC GAG GCC GGC GTC GCG TCA CAG

Met Ser Glu Ala Gly Val Ala Ser Gln

5

10

(D) TOPOLOGY: linear

15

5

25

(D) TOPOLOGY: linear

35

(x1) SEQUENCE DESCRIPTION: SEQ ID NO:32:

27

1 5

5

5

(A) LENGTH: 26 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

10

(11) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:38:

15

GGGGGATCCG TTCTGCCTCC CCGGAC

26

(2) INFORMATION FOR SEQ ID NO:39:

20

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 5312 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

25

(ii) MOLECULE TYPE: DNA

(ix) FEATURE:

(A) NAME/KEY: CDS

30

(B) LOCATION: 37..5145

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:39:

35

GTTCTGCCTC CCCC GGACGG TAAATATAGG GGAACA ATG TAC GCG AAA GCG ACA

54

Met Tyr Ala Lys Ala Thr

	GAA GTC CAG AGG CGC CAC GGC TCC AGC ATT GAG CTG CGC ATC ACT CGC	1014
	Glu Val Gln Arg Arg His Gly Ser Ser Ile Glu Leu Arg Ile Thr Arg	
	315 320 325	
5	GCG CCA CCT GGA GAC CGC ATG CTG GCC GTC GTC CCA AGG ACG TCC CAA	1062
	Ala Pro Pro Gly Asp Arg Met Leu Ala Val Val Pro Arg Thr Ser Gln	
	330 335 340	
	GCG CTC TGC AGA ATC CCA AAC ATC TTT TAT TAC GCC GAC GCG TCG GGC	1110
10	Gly Leu Cys Arg Ile Pro Asn Ile Phe Tyr Tyr Ala Asp Ala Ser Gly	
	345 350 355	
	ACT GAG CAT AAG ACC ATC CTT ACG TCA CAG CAC AAA GTC AAC ATG CTG	1158
	Thr Glu His Lys Thr Ile Leu Thr Ser Gln His Lys Val Asn Met Leu	
15	360 365 370	
	CTC AAT TTT ATG CAA ACG CGT CCT GAG AAG GAA CTA GTC GAC ATG ACC	1206
	Leu Asn Phe Met Gln Thr Arg Pro Glu Lys Glu Leu Val Asp Met Thr	
	375 380 385 390	
20	GTC TTG ATG TCG TTC GCG CGC GCT AGG CTG CGC GCG ATC GTG GTC GCC	1254
	Val Leu Met Ser Phe Ala Arg Ala Arg Leu Arg Ala Ile Val Val Ala	
	395 400 405	
25	TCA GAA GTC ACC GAG AGC TCC TGG AAC ATC TCA CCG GCT GAC CTG GTC	1302
	Ser Glu Val Thr Glu Ser Ser Trp Asn Ile Ser Pro Ala Asp Leu Val	
	410 415 420	
	CGC ACT GTC GTG TCT CTT TAC GTC CTC CAC ATC ATC GAG CGC CGA AGG	1350
30	Arg Thr Val Val Ser Leu Tyr Val Leu His Ile Ile Glu Arg Arg Arg	
	425 430 435	
	GCT GCG GTC GCT GTC AAG ACC GCC AAG GAC GAC GTC TTT GGA GAG ACT	1398
	Ala Ala Val Ala Val Lys Thr Ala Lys Asp Asp Val Phe Gly Glu Thr	
35	440 445 450	

	TCG TTC TGG GAG AGT CTC AAG CAC GTC TTG GGC TCC TGT TGC GGT CTG	1446
	Ser Phe Trp Glu Ser Leu Lys His Val Leu Gly Ser Cys Cys Gly Leu	
	455 460 465 470	
5	CGC AAC CTC AAA GGC ACC GAC GTC GTC TTT ACT AAG CGC GTC GTC GAT	1494
	Arg Asn Leu Lys Gly Thr Asp Val Val Phe Thr Lys Arg Val Val Asp	
	475 480 485	
	AAG TAC CGA GTC CAC TCG CTC GGA GAC ATA ATC TGC GAC GTC CGC CTG	1542
10	Lys Tyr Arg Val His Ser Leu Gly Asp Ile Ile Cys Asp Val Arg Leu	
	490 495 500	
	TCC CCT GAA CAG GTC GGC TTC CTG CCG TCC CGC GTA CCA CCT GCC CGC	1590
	Ser Pro Glu Gln Val Gly Phe Leu Pro Ser Arg Val Pro Pro Ala Arg	
15	505 510 515	
	GTC TTT CAC GAC AGG GAA GAG CTT GAG GTC CTT CGC GAA GCT GGC TGC	1638
	Val Phe His Asp Arg Glu Glu Leu Glu Val Leu Arg Glu Ala Gly Cys	
	520 525 530	
20	TAC AAC GAA CGT CCG GTA CCT TCC ACT CCT CCT GTG GAG GAG CCC CAA	1686
	Tyr Asn Glu Arg Pro Val Pro Ser Thr Pro Pro Val Glu Glu Pro Gln	
	535 540 545 550	
25	GGT TTC GAC GCC GAC TTG TGG CAC GCG ACC GCG GCC TCA CTC CCC GAG	1734
	Gly Phe Asp Ala Asp Leu Trp His Ala Thr Ala Ala Ser Leu Pro Glu	
	555 560 565	
	TAC CGC GCC ACC TTG CAG GCA GGT CTC AAC ACC GAC GTC AAG CAG CTC	1782
30	Tyr Arg Ala Thr Leu Gln Ala Gly Leu Asn Thr Asp Val Lys Gln Leu	
	570 575 580	
	AAG ATC ACC CTC GAG AAC GCC CTC AAG ACC ATC GAC GGG CTC ACC CTC	1830
	Lys Ile Thr Leu Glu Asn Ala Leu Lys Thr Ile Asp Gly Leu Thr Leu	
35	585 590 595	

	TCC CCA GTC AGA GGC CTC GAG ATG TAC GAG GGC CCG CCA GGC AGC GGC	1878
	Ser Pro Val Arg Gly Leu Glu Met Tyr Glu Gly Pro Pro Gly Ser Gly	
	600 605 610	
5	AAG ACG GGC ACC CTC ATC GCC GCC CTT GAG GCC GCG GGC GGT AAA GCA	1926
	Lys Thr Gly Thr Leu Ile Ala Ala Leu Glu Ala Ala Gly Gly Lys Ala	
	615 620 625 630	
	CTT TAC GTG GCA CCC ACC AGA GAA CTG AGA GAG GCT ATG GAC CGG CGG	1974
10	Leu Tyr Val Ala Pro Thr Arg Glu Leu Arg Glu Ala Met Asp Arg Arg	
	635 640 645	
	ATC AAA CCG CCG TCC GCC TCG GCT ACG CAA CAT GTC GCC CTT GCG ATT	2022
15	Ile Lys Pro Pro Ser Ala Ser Ala Thr Gln His Val Ala Leu Ala Ile	
	650 655 660	
	CTC CGT CGT GCC ACC GCC GAG GGC GCC CCT TTC GCT ACC GTG GTT ATC	2070
	Leu Arg Arg Ala Thr Ala Glu Gly Ala Pro Phe Ala Thr Val Val Ile	
	665 670 675	
20	GAC GAG TGC TTC ATG TTC CCG CTC GTG TAC GTC GCG ATC GTG CAC GCC	2118
	Asp Glu Cys Phe Met Phe Pro Leu Val Tyr Val Ala Ile Val His Ala	
	680 685 690	
25	TTG TCC CCG AGC TCA CGA ATA GTC CTT GTA GGG GAC GTC CAC CAA ATC	2166
	Leu Ser Pro Ser Ser Arg Ile Val Leu Val Gly Asp Val His Gln Ile	
	695 700 705 710	
	GGG TTT ATA GAC TTC CAA GGC ACA AGC GCG AAC ATG CCG CTC GTT CGC	2214
30	Gly Phe Ile Asp Phe Gln Gly Thr Ser Ala Asn Met Pro Leu Val Arg	
	715 720 725	
	GAC GTC GTT AAG CAG TGC CGT CGG CGC ACT TTC AAC CAA ACC AAG CGC	2262
	Asp Val Val Lys Gln Cys Arg Arg Arg Thr Phe Asn Gln Thr Lys Arg	
35	730 735 740	

TGT CCG GCC GAC GTC GTT GCC ACC ACG TTT TTC CAG AGC TTG TAC CCC 2310
 Cys Pro Ala Asp Val Val Ala Thr Thr Phe Phe Gln Ser Leu Tyr Pro
 745 750 755

5 GGG TGC ACA ACC ACC TCA GGG TGC GTC GCA TCC ATC AGC CAC GTC GCC 2358
 Gly Cys Thr Thr Thr Ser Gly Cys Val Ala Ser Ile Ser His Val Ala
 760 765 770

10 CCA GAC TAC CGC AAC AGC CAG GCG CAA ACG CTC TGC TTC ACG CAG GAG 2406
 Pro Asp Tyr Arg Asn Ser Gln Ala Gln Thr Leu Cys Phe Thr Gln Glu
 775 780 785 790

15 GAA AAG TCG CGC CAC GGG GCT GAG GGC GCG ATG ACT GTG CAC GAA GCG 2454
 Glu Lys Ser Arg His Gly Ala Glu Gly Ala Met Thr Val His Glu Ala
 795 800 805

20 CAG GGA CGC ACT TTT GCG TCT GTC ATT CTG CAT TAC AAC GGC TCC ACA 2502
 Gln Gly Arg Thr Phe Ala Ser Val Ile Leu His Tyr Asn Gly Ser Thr
 810 815 820

25 GCA GAG CAG AAG CTC CTC GCT GAG AAG TCG CAC CTT CTA GTC GGC ATC 2550
 Ala Glu Gln Lys Leu Leu Ala Glu Lys Ser His Leu Leu Val Gly Ile
 825 830 835

35

	GAC TCG CTC CTC TGT GGT AGC CAT TAC CTC CGT TTC GAC GCT AGC CGC	3750
	Asp Ser Leu Leu Cys Gly Ser His Tyr Leu Arg Phe Asp Ala Ser Arg	
	1225 1230 1235	
5	CTT CAC ATG GGC GAA CGT TAC AAG ACC AAA CAT TTG AAG GTC GAG GTG	3798
	Leu His Met Gly Glu Arg Tyr Lys Thr Lys His Leu Lys Val Glu Val	
	1240 1245 1250	
	CAG AAA ATC GTG CCG TAC ATC GGA CTC CTC GTC TCC GCT GAG CAG GTC	3846
10	Gln Lys Ile Val Pro Tyr Ile Gly Leu Leu Val Ser Ala Glu Gln Val	
	1255 1260 1265 1270	
	GTC CTC GAC CCT GTC AGG AGC GCT CTC AAG ATA TTT GGG CGC TGC TAC	3894
	Val Leu Asp Pro Val Arg Ser Ala Leu Lys Ile Phe Gly Arg Cys Tyr	
15	1275 1280 1285	
	ACA AGC GAA CTC CTT TAC TCC AAG TAC GTG GAG GCT GTG AGA GAC ATC	3942
	Thr Ser Glu Leu Leu Tyr Ser Lys Tyr Val Glu Ala Val Arg Asp Ile	
	1290 1295 1300	
20	ACC AAG GGC TGG AGT GAC GCC CGC TAC CAC AGC CTC CTG TGC CAC ATG	3990
	Thr Lys Gly Trp Ser Asp Ala Arg Tyr His Ser Leu Leu Cys His Met	
	1305 1310 1315	
25	TCA GCA TGC TAC TAC AAT TAC GCG CCG GAG TCT GCG GCG TAC ATC ATC	4038
	Ser Ala Cys Tyr Tyr Asn Tyr Ala Pro Glu Ser Ala Ala Tyr Ile Ile	
	1320 1325 1330	
	GAC GCT GTT GTT CGC TTT GGG CGC GGC GAC TTC CCG TTT GAA CAA CTG	4086
30	Asp Ala Val Val Arg Phe Gly Arg Gly Asp Phe Pro Phe Glu Gln Leu	
	1335 1340 1345 1350	
	CGC GTG GTG CGT GCC CAT GTG CAG GCA CCC GAC GCT TAC AGC AGC ACG	4134
	Arg Val Val Arg Ala His Val Gln Ala Pro Asp Ala Tyr Ser Ser Thr	
35	1355 1360 1365	

	TAT CCG GCT AAC GTG CGC GCA TCG TGC CTT GAC CAC GTC TTC GAG CCC	4182
	Tyr Pro Ala Asn Val Arg Ala Ser Cys Leu Asp His Val Phe Glu Pro	
	1370 1375 1380	
5	CGC CAG GCC GCC GCC CCG GCA GGT TTC GTT GCG ACA TGT GCG AAG CCG	4230
	Arg Gln Ala Ala Ala Pro Ala Gly Phe Val Ala Thr Cys Ala Lys Pro	
	1385 1390 1395	
10	GAA ACG CCT TCT TCA CTT ACC GCG AAA GCT GGT GTT TCT GCG ACT ACA	4278
	Glu Thr Pro Ser Ser Leu Thr Ala Lys Ala Gly Val Ser Ala Thr Thr	
	1400 1405 1410	
15	AGC CAC GTT GCG ACT GGG ACT GCG CCC CCG GAG TCT CCA TGG GAT GCA	4326
	Ser His Val Ala Thr Gly Thr Ala Pro Pro Glu Ser Pro Trp Asp Ala	
	1415 1420 1425 1430	
20	CCT GCA GCC AAC AGC TTT TCG GAG TTA TTG ACA CCG GAG ACC CCG TCC	4374
	Pro Ala Ala Asn Ser Phe Ser Glu Leu Leu Thr Pro Glu Thr Pro Ser	
	1435 1440 1445	
25	ACA TCA TCC TCG CCG TCA TCG TCT TCA TCG GAC TCC TCT ACA TCG TGT	4422
	Thr Ser Ser Ser Pro Ser Ser Ser Ser Ser Asp Ser Ser Thr Ser Cys	
	1450 1455 1460	
30	GGA AGG TCG CTC AGT GGT GGA GAC ACC GCA AGG ACC ACA GAA GAC TTG	4470
	Gly Arg Ser Leu Ser Gly Gly Asp Thr Ala Arg Thr Thr Glu Asp Leu	
	1465 1470 1475	
35	AAC AGC AGA AAG CCG CCT TCG CAA GAC AGG CAA TCA CGC TCG TCT GAA	4518
	Asn Ser Arg Lys Pro Pro Ser Gln Asp Arg Gln Ser Arg Ser Ser Glu	
	1480 1485 1490	
	TGT CTG GAC AGA AGC GGA GAA AGG ACA GGC AGT TCG TTA ACT GCC CCC	4566
	Cys Leu Asp Arg Ser Gly Glu Arg Thr Gly Ser Ser Leu Thr Ala Pro	
	1495 1500 1505 1510	

	ACT GCT CCG AGC CCC TCA TTC TCA TTT TCG GAA AGA GCT CGA CTG GCG	4614
	Thr Ala Pro Ser Pro Ser Phe Ser Phe Ser Glu Arg Ala Arg Leu Ala	
	1515 1520 1525	
5	ACC GGG CCG ACT GTC GCC GCT GCG ACA TCA CCT TCG GCA ACC CCA TCC	4662
	Thr Gly Pro Thr Val Ala Ala Ala Thr Ser Pro Ser Ala Thr Pro Ser	
	1530 1535 1540	
10	TGC GCC ACG GAC CAG GTT GCC GCG AGG ACC ACG CCG GAC TTT GCG CCT	4710
	Cys Ala Thr Asp Gln Val Ala Ala Arg Thr Thr Pro Asp Phe Ala Pro	
	1545 1550 1555	
15	TTC CTG GGT TCC CAG TCT GCC CGT GCT GTC TCG AAG CCG TAC CGG CCC	4758
	Phe Leu Gly Ser Gln Ser Ala Arg Ala Val Ser Lys Pro Tyr Arg Pro	
	1560 1565 1570	
20	CCC ACG ACT GCC CGT TGG AAA GAA GTC ACC CCG CTC CAC GCG TGG AAG	4806
	Pro Thr Thr Ala Arg Trp Lys Glu Val Thr Pro Leu His Ala Trp Lys	
	1575 1580 1585 1590	
25	GGC GTG ACC GGA GAC CGA CCG GAA GTC AGG GAG GAC CCG GAG ACA GCG	4854
	Gly Val Thr Gly Asp Arg Pro Glu Val Arg Glu Asp Pro Glu Thr Ala	
	1595 1600 1605	
30	GCG GTC GTC CAG GCT CTG ATC AGC GGC CGT TAT CCT CAG AAG ACG AAG	4902
	Ala Val Val Gln Ala Leu Ile Ser Gly Arg Tyr Pro Gln Lys Thr Lys	
	1610 1615 1620	
35	CTT TCC TCC GAC GCA TCC AAA GGC TAC TCA AGA ACT AAG GGA TGC TCA	4950
	Leu Ser Ser Asp Ala Ser Lys Gly Tyr Ser Arg Thr Lys Gly Cys Ser	
	1625 1630 1635	
40	CAA TCC ACC TCT TTT CCT GCC CCG AGT GCG GAT TAC CAG GCC CGC GAC	4998
	Gln Ser Thr Ser Phe Pro Ala Pro Ser Ala Asp Tyr Gln Ala Arg Asp	
	1640 1645 1650	

35

35

35

35

35

35

1230

Figure 1 consists of six panels (a-f) showing the effect of the 1997-1998 El Niño on the growth of 10-year-old trees in a tropical forest. The panels are arranged in a 3x2 grid. The x-axis for all panels is 'Growth rate (cm year⁻¹)' and the y-axis is 'Number of trees'. The panels show the distribution of growth rates for different tree species and sizes. Panel (a) shows growth rates for 10-year-old trees in 1997-1998 compared to 1996-1997. Panel (b) shows growth rates for 10-year-old trees in 1997-1998 compared to 1996-1997. Panel (c) shows growth rates for 10-year-old trees in 1997-1998 compared to 1996-1997. Panel (d) shows growth rates for 10-year-old trees in 1997-1998 compared to 1996-1997. Panel (e) shows growth rates for 10-year-old trees in 1997-1998 compared to 1996-1997. Panel (f) shows growth rates for 10-year-old trees in 1997-1998 compared to 1996-1997.

5

10

15

20

25

30

35

Gly Val Ser Ala Thr Thr Ser His Val Ala Thr Gly Thr Ala Pro Pro
1410 1415 1420

35

5

10

Asp Tyr Gln Ala Arg Asp Cys Gln Thr Val Arg Val Cys Arg Ala Ala
1650 1655 1660

Ala Glu Met Ala Arg Ser Cys Ile His Glu Pro Leu Ala Ser Ser Ala
1665 1670 1675 1680

5

10

(1) SEQUENCE CHARACTERISTICS:

- 15

20

(A) NAME/KEY: CDS

- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:41:

25

GTTCTGCCCTC CCCCGGACGG TAAATATAGG GGAACAATGT ACGCGAAAGC GACAGACGTG 60

GC GCGTGTCT ACGCCGCGGC AGATGTCGCC TACGCGAACG TACTGCAGCA GAGAGCAGTC 120

30

AAGTTGGACT TCGCCCCGCC ACTGAAGGCA CTAGAAACCC TCCACAGACT GTACTATCCG 180

CTGCGCTTCA AAGGGGGCAC TTTACCCCCG ACACAACACC CGATCCTGGC CGGGCACCAA 240

CGTGTGCGAG AAGAGGTTCT GCACAATTC GCCAGGGGAC GTAGCACAGT GCTCGAGATA 300

35

GGGCCGTCTC TGCACAGCGC ACTTAAGCTA CATGGGGCAC CGAACGCCCC CGTCGCAGAC 360

35

CGAGTCCACT CGCTCGGAGA CATAATCTGC GACGTCCGCC TGTCCCCTGA ACAGGTCGGC 1560

5 TTCCTGCCGT CCCGCGTACC ACCTGCCCGC GTCTTTCACG ACAGGGAAGA GCTTGAGGTC 1620

CTTCGCGAAG CTGGCTGCTA CAACGAACGT CCGGTACCTT CCACTCCTCC TGTGGAGGAG 1680

CCCCAAGGTT TCGACGCCGA CTTGTGGCAC GCGACCGCGG CCTCACTCCC CGAGTACCGC 1740

10

GCCACCTTGC AGGCAGGTCT CAACACCGAC GTCAAGCAGC TCAAGATCAC CCTCGAGAAC 1800

	GCCCTCAAGA CCATCGACGG GCTCACCCCTC TCCCCAGTCA GAGGCCTCGA GATGTACGAG	1860
	GGCCCGCCAG GCAGCGGCAA GACGGGCACC CTCATCGCCG COCTTGAGGC CGCGGGCGGT	1920
5	AAAGCACTTT ACGTGGCACC CACCAGAGAA CTGAGAGAGG CTATGGACCG GCGGATCAAA	1980
	CCGCGTCCG CCTCGGCTAC GCAACATGTC GCCCTTGCGA TTCTCCGTCTG TGCCACCGCC	2040
	GAGGGCGCCC CTTTCGCTAC CGTGGTTATC GACGAGTGCT TCATGTTCCC GCTCGTGTAC	2100
10	GTGCGATCG TGACACGCTT GTCCCCGAGC TCACGAATAG TCCTTGTAGG GGACGTCCAC	2160
	CAAATCGGGT TTATAGACTT CCAAGGCACA AGCGCGAACA TGCCGCTCGT TCGCGACGTC	2220
15	GTTAAGCAGT GCGTCGGCG CACTTTCAAC CAAACCAAGC GCTGTCCGGC CGACGTCGTT	2280
	GCCACCACGT TTTTCAGAG CTTGTACCCC GGETGCACAA CCACCTCAGG GTGCGTCGCA	2340
	TCCATCAGCC ACGTCGCCCC AGACTACCGC AACAGCCAGG CGCAAACGCT CTGCTTCACG	2400
20	CAGGAGGAAA AGTCGCGCCA CGGGGCTGAG GCGCGATGA CTGTGCACGA AGCGCAGGGA	2460
	CGCACTTTTG CGTCTGTCAT TCTGCATTAC AACGGCTCCA CAGCAGAGCA GAAGCTCCTC	2520
25	GCTGAGAAGT CGCACCTTCT AGTCGGCATC ACGCGCCACA CCAACCACCT GTACATCCGC	2580
	GACCCGACAG GTGACATTGA GAGACAACTC AACCATAGCG CGAAAGCCGA GGTGTTTACA	2640
	GACATCCCTG CACCCCTGGA GATCACGACT GTCAAACCGA GTGAAGAGGT GCAGCGCAAC	2700
30	GAAGTGATGG CAACGATACC CCCGCAGAGT GCCACGCCGC ACGGAGCAAT CCATCTGCTC	2760
	CGCAAGAACT TCGGGGACCA ACCCGACTGT GGCTGTGTCG CTTTGGCGAA GACCGGCTAC	2820
35	GAGGTGTTTG GCGGTGCTGC CAAAATCAAC GTAGAGCTTG CCGAACCCGA CGCGACCCCG	2880

CCGCGTCCG CCTCGGCTAC GCAACATGTC GCCCTTGCGA TTCTCCGTCTG TGCCACCGCC

	AAGCCGCATA GGGCGTCCA GGAAGGGETA CAGTGGGTCA AGGTCACCAA CGCGTCTAAC	2940
	AAACACCAGG CGCTCCAGAC GCTGTTGTCC CGCTACACCA AGCGAAGCGC TGACCTGCCG	3000
5	CTACACGAAG CTAAGGAGGA CGTCAAACGC ATGCTAAACT CGCTTGACCG ACATTGGGAC	3060
	TGGACTGTCA CTGAAGACGC CCGTGACCGA GCTGTCTTCG AGACCCAGCT CAAGTTCACC	3120
	CAACGCGGCG GCACCGTCGA AGACCTGCTG GAGCCAGACG ACCCTACAT CCGTGACATA	3180
10	GACTTCCTTA TGAAGACTCA GCAGAAAGTG TCGCCCAAGC CGATCAATAC GGGCAAGGTC	3240
	GGGCAGGGGA TCGCCGCTCA CTCAAAGTCT CTCAACTTCG TCCTCGCCGC TTGGATACGC	3300
15	ATACTCGAGG AGATACTCCG TACCGGGAGC CGCACGGTCC GGTACAGCAA CGGTCTCCCC	3360
	GACGAAGAAG AGGCCATGCT GCTCGAAGCG AAGATCAATC AAGTCCCACA CGCCACGTTC	3420
	GTCTCGGCGG ACTGGACCGA GTTTGACACC GCCCACAATA ACACGAGTGA GCTGCTCTTC	3480
20	GCGGCCCTTT TAGAGCGCAT CGGCACGCCT GCAGCTGCCG TTAATCTATT CAGAGAACGG	3540
	TGTGGGAAAC GCACCTTGGC AGCGAAGGGT CTAGGCTCCG TTGAAGTCGA CGGTCTGCTC	3600
25	GACTCCGGCG CAGCTTGGAC GCCTTGCCGC AACACCATCT TCTCTGCCGC CGTCATGCTC	3660
	ACGCTCTTCC GCGGCGTCAA GTTCGCAGCT TTCAAAGGCG ACGACTCGCT CCTCTGTGCT	3720
	AGCCATTACC TCCGTTTCGA CGCTAGCCGC CTTACATGG GCGAACGTTA CAAGACCAA	3780
30	CATTTGAAGG TCGAGGTGCA GAAATCGTG CCGTACATCG GACTCCTCGT CTCGCTGAG	3840
	CAGGTCGTCC TCGACCTGT CAGGAGCGCT CTCAGATAT TTGGCGCTG CTACACAAGC	3900
35	GAACCTCCTTT ACTCCAAGTA CGTGGAGGCT GTGAGAGACA TCACCAAGGG CTGGAGTGAC	3960

TCTGCGGCGT ACATCATCGA CGCTGTTGTT CGCTTTGGGC GCGGCGACTT CCCGTTTGAA 4080

5 CAACTGCGCG TGGTGGGTGC CCATGTGCAG GCACCGGACG CTTACAGCAG CACGTATCCG 4140

GCTAACGTGC GCGCATCGTG CCTTGACCAC GTCTTCGAGC CCCGCCAGGC CGCCGCCCCG 4200

GCAGGTTTCG TTGCGAC ATG TGC GAA GCC GGA AAC GCC TTC TTC ACT TAC 4250

10

Met Cys Glu Ala Gly Asn Ala Phe Phe Thr Tyr

1 5 10

	CGC GAA AGC TGG TGT TTC TGC GAC TAC AAG CCA CGT TGC GAC TGG GAC	4298
	Arg Glu Ser Trp Cys Phe Cys Asp Tyr Lys Pro Arg Cys Asp Trp Asp	
	15 20 25	
5	TGC GCC CCC GGA GTC TCC ATG GGA TGC ACC TGC AGC CAA CAG CTT TTC	4346
	Cys Ala Pro Gly Val Ser Met Gly Cys Thr Cys Ser Gln Gln Leu Phe	
	30 35 40	
	GGA GTT ATT GAC ACC GGA GAC CCC GTC CAC ATC ATC CTC GCC GTC ATC	4394
10	Gly Val Ile Asp Thr Gly Asp Pro Val His Ile Ile Leu Ala Val Ile	
	45 50 55	
	GTC TTC ATC GGA CTC CTC TAC ATC GTG TGG AAG GTC GCT CAG TGG TGG	4442
	Val Phe Ile Gly Leu Leu Tyr Ile Val Trp Lys Val Ala Gln Trp Trp	
15	60 65 70 75	
	AGA CAC CGC AAG GAC CAC AGA AGA CTT GAA CAG CAG AAA GCC GCC TTC	4490
	Arg His Arg Lys Asp His Arg Arg Leu Glu Gln Gln Lys Ala Ala Phe	
	80 85 90	
20	GCA AGA CAG GCA ATC ACG CTC GTC TGAATGTC TGGACAGAAG CGGAGAAAGG	4542
	Ala Arg Gln Ala Ile Thr Leu Val	
	95	
25	ACAGGCAGTT CGTTAACTGC CCCACTGCT CCGAGCCCTT CATTCTCATT TTCGGAAAGA	4602
	GCTCGACTGG CGACCGGGCC GACTGTGCC GCTGCGACAT CACCTTCGGC AACCCCATCC	4662
	TGCGCCACGG ACCAGGTTGC CGCGAGGACC ACGCCGGACT TTGCGCCTTT CCTGGGTTCC	4722
30	CAGTCTGCC GTGCTGTCTC GAAGCCGTAC CGGCCCCCA CGACTGCCCG TTGGAAGAA	4782
	GTCACCCCGC TCCACGCGTG GAAGGCGGTG ACCGGAGACC GACCGGAAGT CAGGGAGGAC	4842
35	CCGGAGACAG CGGCGGTCGT CCAGGCTCTG ATCAGCGGCC GTTATCCTCA GAAGACGAAG	4902

(2) INFORMATION FOR SEQ ID NO:42:

5 (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 99 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

10 (ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:42:

15 Met Cys Glu Ala Gly Asn Ala Phe Phe Thr Tyr Arg Glu Ser Trp Cys
1 5 10 15

Phe Cys Asp Tyr Lys Pro Arg Cys Asp Trp Asp Cys Ala Pro Gly Val
20 25 30

20 Ser Met Gly Cys Thr Cys Ser Gln Gln Leu Phe Gly Val Ile Asp Thr
35 40 45

Gly Asp Pro Val His Ile Ile Leu Ala Val Ile Val Phe Ile Gly Leu
25 50 55 60

Leu Tyr Ile Val Trp Lys Val Ala Gln Trp Trp Arg His Arg Lys Asp
65 70 75 80

30 His Arg Arg Leu Glu Gln Gln Lys Ala Ala Phe Ala Arg Gln Ala Ile
85 90 95

Thr Leu Val

35

5

10

15

(D) TOPOLOGY: linear

(B) LOCATION: 4518..4937

20

25

30

35

GTTGGAATTG CCAATCACTC CCTCTATGAC GTGACCCTAG AGGAGCTGGC CAATGCGTTT 600

35

	CAATCACGCT CGTCTGA ATG TCT GGA CAG AAG CGG AGA AAG GAC AGG CAG	4550
	Met Ser Gly Gln Lys Arg Arg Lys Asp Arg Gln	
	1 5 10	
5	TTC GTT AAC TGC CCC CAC TGC TCC GAG CCC CTC ATT CTC ATT TTC GGA Phe Val Asn Cys Pro His Cys Ser Glu Pro Leu Ile Leu Ile Phe Gly	4598
	15 20 25	
	AAG AGC TCG ACT GGC GAC CGG GCC GAC TGT CGC CGC TGC GAC ATC ACC	4646
10	Lys Ser Ser Thr Gly Asp Arg Ala Asp Cys Arg Arg Cys Asp Ile Thr	
	30 35 40	
	TTC GGC AAC CCC ATC CTG CGC CAC GGA CCA GGT TGC CGC GAG GAC CAC	4694
	Phe Gly Asn Pro Ile Leu Arg His Gly Pro Gly Cys Arg Glu Asp His	
15	45 50 55	
	GCC GGA CTT TGC GCC TTT CCT GGG TTC CCA GTC TGC CCG TGC TGT CTC	4742
	Ala Gly Leu Cys Ala Phe Pro Gly Phe Pro Val Cys Pro Cys Cys Leu	
	60 65 70 75	
20	GAA GCC GTA CCG GCC CCC CAC GAC TGC CCG TTG GAA AGA AGT CAC CCC	4790
	Glu Ala Val Pro Ala Pro His Asp Cys Pro Leu Glu Arg Ser His Pro	
	80 85 90	
25	GCT CCA CGC GTG GAA GGG CGT GAC CGG AGA CCG ACC GGA AGT CAG GGA	4838
	Ala Pro Arg Val Glu Gly Arg Asp Arg Arg Pro Thr Gly Ser Gln Gly	
	95 100 105	
	GGA CCC GGA GAC AGC GGC GGT CGT CCA GGC TCT GAT CAG CGG CCG TTA	4886
30	Gly Pro Gly Asp Ser Gly Gly Arg Pro Gly Ser Asp Gln Arg Pro Leu	
	110 115 120	
	TCC TCA GAA GAC GAA GCT TTC CTC CGA CGC ATC CAA AGG CTA CTC AAG	4934
	Ser Ser Glu Asp Glu Ala Phe Leu Arg Arg Ile Gln Arg Leu Leu Lys	
35	125 130 135	

140

GTATTCACGA GCCGTTGGCT TCATCTGCCG CCAGTGCCGA CTTGAAGCGC ATACGCTCTA 5107

10

TCCTTGTTCTG TAAACAAGGT GGTCCCTCCC ATTGAGGTAA AGACTCTGGT GAGTCCTCAA 5227

TAGTACGGCG CCCCTGGGA TACCA 5312

(2) INFORMATION FOR SEQ ID NO:44:

10

(B) TYPE: amino acid

(D) TOPOLOGY: linear

15

(x1) SEQUENCE DESCRIPTION: SEQ ID NO:44:

20

1 5 10 15

20 25 30

25

Asp Arg Ala Asp Cys Arg Arg Cys Asp Ile Thr Phe Gly Asn Pro Ile

35 40 45

50 55 60

30

Phe Pro Gly Phe Pro Val Cys Pro Cys Cys Leu Glu Ala Val Pro Ala

65 70 75 80

Pro His Asp Cys Pro Leu Glu Arg Ser His Pro Ala Pro Arg Val Glu

35

85 90 95

5

Gly Gly Arg Pro Gly Ser Asp Gln Arg Pro Leu Ser Ser Glu Asp Glu

115 120 125

10

130 135 140

(2) INFORMATION FOR SEQ ID NO:45:

(i) SEQUENCE CHARACTERISTICS:

15

(A) LENGTH: 5312 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

20

(ii) MOLECULE TYPE: DNA

(1X) FEATURE:

(A) NAME/KEY: CDS

(B) LOCATION: 4944..5162

25

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:45:

30

GTTCTGCCTC CCCC GGACGG TAAATATAGG GGAACAATGT ACGCGAAAGC GACAGACGTG 60

GCGCGTGTCT ACGCCGCGGC AGATGTCGCC TACGCGAACG TACTGCAGCA GAGAGCAGTC 120

AAGTTGGACT TCGCCCCGCC ACTGAAGGCA CTAGAAACCC TCCACAGACT GTACTATCCG 180

35

CTGCGCTTCA AAGGGGGGCAC TTTACCCCCG ACACAACACC CGATCCTGGC CGGGCACCAA 240

5

10

	GAGAACCACG GACTTCACAT GGTCGCGCG TTCATGCACA TGCCAGAAGA GCTGCTCTAC	660
	ATGGACAACG TGGTTAATGC CGAGCTCGGC TACCGCTTCC ACGTTATTGA AGAGCCTATG	720
5	GCTGTGAAGG ACTGCGCATT CCAGGGGGGG GACCTCCGTC TCCACTTCCC TGAGTTGGAC	780
	TTCATCAACG AGAGCCAAGA GCGGCGCATC GAGAGSCTGG CCGCCCGCGG CTCCTACTCC	840
	AGACGCGCCG TCATTTTCTC CGGCGACGAC GACTGGGGTG ATGCGTACTT ACACGACTTC	900
10	CACACATGGC TCGCCTACCT ACTGGTGAGG AACTACCCCA CTCCGTTTGG TTTCTCACTC	960
	CATATAGAAG TCCAGAGGCG CCACGGCTCC AGCATTGAGC TGCGCATCAC TCGCGCGCCA	1020
15	CCTGGAGACC GCATGCTGGC CGTCGTCCCA AGGACGTCCC AAGGCTCTG CAGAATCCCA	1080
	AACATCTTTT ATTACGCCGA CGCGTCGGGC ACTGAGCATA AGACCATCCT TACGTCACAG	1140
	CACAAAGTCA ACATGCTGCT CAATTTTATG CAAACGGGTC CTGAGAAGGA ACTAGTCGAC	1200
20	ATGACCGTCT TGATGTCGTT CGCGCGCGCT AGGCTGCGCG CGATCGTGGT CGCCTCAGAA	1260
	GTCACCGAGA GCTCCTGGAA CATCTCACCG GCTGACCTGG TCCGCACTGT CGTGTCTCTT	1320
25	TACGTCTCTC ACATCATCGA GCGCCGAAGG GCTGCGGTG CTGTCAAGAC CGCCAAGGAC	1380
	GACGTCTTTG GAGAGACTTC GTTCTGGGAG AGTCTCAAGC ACGTCTTGGG CTCCTGTTGC	1440
	GGTCTGCGCA ACCTCAAAGG CACCGACGTC GTCTTTACTA AGCGCGTCGT CGATAAGTAC	1500
30	CGAGTCCACT CGCTCGGAGA CATAATCTGC GACGTCCGCC TGTCCCCTGA ACAGGTGCGC	1560
	TTCTTGCCGT CCCGCGTACC ACCTGCCCCG GTCTTTCACG ACAGGGAAGA GCTTGAGGTC	1620
35	CTTCGCGAAG CTGGCTGCTA CAACGAACGT CCGGTACCTT CCACTCTCTC TGTGGAGGAG	1680

	CCCCAAGGTT TCGACGCCGA CTTGTGGCAC GCGACCGCGG CCTCACTCCC CGAGTACCGC	1740
	GCCACCTTGC AGGCAGGTCT CAACACCGAC GTCAAGCAGC TCAAGATCAC CCTCGAGAAC	1800
5	GCCCTCAAGA CCATCGACGG GCTCACCTC TCCCCAGTCA GAGGCCTCGA GATGTACGAG	1860
	GGCCCGCCAG GCAGCGGCAA GACGGGCACC CTCATCGCCG CCCTTGAGGC CGCGGGCGGT	1920
	AAAGCACTTT ACGTGGCACC CACCAGAGAA CTGAGAGAGG CTATGGACCG CGGGATCAAA	1980
10	CCGCGGTCCG CCTCGGCTAC GCAACATGTC GCCCTTGCGA TTCTCCGTCG TGCCACCGCC	2040
	GAGGGCGCCC CTTTCGCTAC CGTGGTTATC GACGAGTGCT TCATGTTCCC GTCGTGTAC	2100
15	GTGCGGATCG TGCACGCCTT GTCCCGGAGC TCACGAATAG TCCTTGTAGG GGACGTCCAC	2160
	CAATCGGGT TTATAGACTT CCAAGGCACA AGCGGAACA TGCCGCTGCT TCGCGACGTC	2220
	GTTAAGCAGT GCCGTCGGCG CACTTTC AACAAACGAG GCTGTCCGGC CGACGTCGTT	2280
20	GCCACCACGT TTTTCAGAG CTTGTACCCC GGGTGCACAA CCACCTCAGG GTGCGTCGCA	2340
	TCCATCAGCC ACGTCGCCCC AGACTACCGC AACAGCCAGG CGAAACGCT CTGCTTCACG	2400
25	CAGGAGGAAA AGTCGCGCCA CGGGGCTGAG GGCAGCATGA CTGTGCACGA AGCGCAGGGA	2460
	CGCACTTTTG CGTCTGTGAT TCTGCATTAC AACGGCTCCA CAGCAGAGCA GAAGCTCCTC	2520
	GCTGAGAAGT CGCACCTTCT AGTCGGCATC ACGCGCCACA CCAACCACCT GTACATCCGC	2580
30	GACCCGACAG GTGACATTGA GAGACAACTC AACCATAGCG CGAAAGCCGA GGTGTTTACA	2640
	GACATCCCTG CACCCCTGGA GATCAGCACT GTCAACCGA GTGAAGAGGT GCAGCGCAAC	2700
35	GAAGTGATGG CAACGATACC CCGCAGAGT GCCACGCCG ACGGAGCAAT CCATCTGCTC	2760

TGGACTGTCA CTGAAGACGC CCGTGACCGA GCTGTCTTCG AGACCCAGCT CAAGTTCACC 3120

	CAACGCGGCG GCACCGTCGA AGACCTGCTG GAGCCAGACG ACCCCTACAT CCGTGACATA	3180
	GACTTCCTTA TGAAGACTCA GCAGAAAGTG TCGCCCAAGC CGATCAATAC GGGCAAGGTC	3240
5	GGGCAGGGGA TCGCCGCTCA CTCAAAGTCT CTCAACTTCG TCCTCGCCGC TTGGATACGC	3300
	ATACTCGAGG AGATACTCCG TACCGGGAGC CGCACGGTCC GGTACAGCAA CGGTCTCCCC	3360
	GACGAAGAAG AGGCCATGCT GCTCGAAGCG AAGATCAATC AAGTCCCACA CGCCACGTTC	3420
10	GTCTCGGCGG ACTGGACCGA GTTTGACACC GCCCACAATA ACACGAGTGA GCTGCTCTTC	3480
	GCCGCCCTTT TAGAGCGCAT CGGCACGCCT GCAGCTGCCG TTAATCTATT CAGAGAACGG	3540
15	TGTGGGAAAC GCACCTTGCG AGCGAAGGGT CTAGGCTCCG TTGAAGTCGA CGGTCTGCTC	3600
	GACTCCGGCG CAGCTTGGAC GCCTTGCCGC AACACCATCT TCTCTGCCGC CGTCATGCTC	3660
	ACGCTCTTTC GCGGCGTCAA GTTCGCAGCT TTCAAAGGCG ACGACTCGCT CCTCTGTGGT	3720
20	AGCCATTACC TCCGTTTCGA CGCTAGCCGC CTTACATGG GCGAACGTTA CAAGACCAA	3780
	CATTTGAAG TCGAGGTGCA GAAAATCGTG CCGTACATCG GACTCCTCGT CTCCGCTGAG	3840
25	CAGGTCGTCC TCGACCTGT CAGGAGCGCT CTCAGATAT TTGGCGCTG CTACACAAGC	3900
	GAACTCCTTT ACTCCAAGTA CGTGGAGGCT GTGAGAGACA TCACCAAGGG CTGGAGTGAC	3960
	GCCCGCTACC ACAGCCTCCT GTGCCACATG TCAGCATGCT ACTACAATTA CGCGCCGGAG	4020
30	TCTGCGGCGT ACATCATCGA CGCTGTIGTT CGCTTTGGGC GCGGCGACTT CCCGTTTGAA	4080
	CAACTGCGCG TGGTGCGTGC CCATGTGCAG GCACCCGACG CTTACAGCAG CACGTATCCG	4140
35	GCTAACGTGC GCGCATCGTG CCTTGACCAC GTCTTCGAGC CCCGCCAGGC CGCCGCCCGG	4200

	GCAGGTTTCG TTGCGACATG TGCGAAGCCG GAAACGCCTT CTTCATTAC CGCGAAAGCT	4260
	GGTGTTCCTG CCACTACAAG CCACGTTGCG ACTGGGACTG CGCCCCCGGA GTCTCCATGG	4320
5	GATGCACCTG CAGCCAACAG CTTTTCGGAG TTATTGACAC CGGAGACCCC GTCCACATCA	4380
	TCTCGCCGT CATCGTCTTC ATCGGACTCC TCTACATCGT GTGGAAGGTC GCTCAGTGGT	4440
	GGAGACACCG CAAGGACCAC AGAAGACTTG AACAGCAGAA AGCCGCCTTC GCAAGACAGG	4500
10	CAATCACGCT CGTCTGAATG TCTGGACAGA AGCGGAGAAA GGACAGGCAG TTCGTTAACT	4560
	GCCCCACTG CTCCGAGCCC CTCATTCTCA TTTCGGAAA GAGCTCGACT GGCACCCGGG	4620
15	CCGACTGTCTG CCGCTGCGAC ATCACCTTCG GCAACCCCAT CCTGCGCCAC GGACCAGGTT	4680
	GCCGCGAGGA CCACGCCGGA CTTTGCGCCT TTCTGGGTT CCCAGTCTGC CCGTGTCTGC	4740
	TCGAAGCCGT ACCGGCCCC CACGACTGCC CGTTGGAAAG AAGTACCCCC GCTCCACGCG	4800
20	TGGAAGGGCG TGACCGGAGA CCGACCGGAA GTCAGGGAGG ACCCGGAGAC AGCGGCGGTC	4860
	GTCCAGGCTC TGATCAGCGG CCGTTATCCT CAGAAGACGA AGCTTTCTCT CGACGCATCC	4920
25	AAAGGCTACT CAAGAACTAA GGG ATG CTC ACA ATC CAC CTC TTT TCC TGC	4970
	Met Leu Thr Ile His Leu Phe Ser Cys	
	1 5	
	CCC GAG TGC GGA TTA CCA GGC CCG CGA CTG CCA GAC AGT CCG AGT CTG	5018
30	Pro Glu Cys Gly Leu Pro Gly Pro Arg Leu Pro Asp Ser Pro Ser Leu	
	10 15 20 25	
	CCG CGC CGC TGC AGA GAT GGC GCG CTC ATG TAT TCA CGA GCC GTT GGC	5066
	Pro Arg Arg Cys Arg Asp Gly Ala Leu Met Tyr Ser Arg Ala Val Gly	
35	30 35 40	

45 50 55

60 65 70

10

GCAACTAGTA CGGCGCCCCC TGGGATACCA 5312

(2) INFORMATION FOR SEQ ID NO:46:

20

(B) TYPE: amino acid

(11) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:46:

30

20 25 30

35 40 45

50 55 60

5 65 70

10

(A) LENGTH: 2478 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(11) MOLECULE TYPE: DNA

(ix) FEATURE:

(A) NAME/KEY: CDS

(B) LOCATION: 283..753

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:47:

GT TTTTCTTT CTTTACCAAG TGTGGTAAAA TTAAACAAA GAAGAAAACC AGGACCGTAA 60

CCCGGCCCTT ACACACCTCG AGTCCGTGAC CACCGGATTA TACGTCGCCC ACCACACGGC 120

GCCTTTTCCG ACCACTCTCG AGAGTCGTTG GGAGTTTCGT CCGTGACCAC CCGGTTGGCA 180

GTCGACAGAC GCTTCCGGAC CACTAGAACC TCCTCGAGCG ACGCACACAC AGCACACACA 240

CCGCGCTTAGC TGCACCTACG GCAGCGTTGA TAGCGCGGAT TT ATG AGC GAG CAC 294

Met Ser Glu His

15

35

	CTCGTCGACC AAGGCTTTTG GATCGGCGGC CAGTACGCTC TCACCCCGAC TAGCCTACCG	1193
	CAGTACGACG TCAGCGAGGC CTACGCTCTG CACACTTTGA CCTTCGCCAG ACCATCCAGC	1253
5	GCCGCTGCAC TCGCGTTTGT GTGGGCAGGT TTGCCACAGG GTGGCACTGC GCCTGCAGGC	1313
	ACTCCAGCCT GGGAGCAGGC ATCCTCGGCT GGCTACCTCA CCTGGCGCCA CAACGGTACT	1373
	ACTTTCCAG CTGGCTCCGT TAGCTACGTT CTCCTGAGG GTTTCGCCCT TGAGCGCTAC	1433
10	GACCCGAACG ACGGCTCTTG GACCGACTTC GCTTCGCAG GAGACACCCT CACTTTCGG	1493
	CAGGTCGCCG TCGACGAGGT CGTTGTGACC AACAAACCCG CCGCGCGCGG CAGCGCCCC	1553
15	ACCTTCACCG TGAGAGTGCC CCCTTCAAAC GCTTACACCA ACACCGTGTT TAGGAACACG	1613
	CTCTTAGAGA CTCGACCCCT CTCTCGTAGG CTCGAACTCC CTATGCCACC TGCTGACTTT	1673
	GGACAGACGG TCGCCAACAA CCGAAGATC GAGCAGTCGC TTCTTAAAGA AACACTTGGC	1733
20	TGCTATTTGG TCCACTCCAA AATGCGAAAC CCCGTTTTC AGCTCAGCC AGCCAGCTCC	1793
	TTTGCGCCG TTCTCTTCAA CAATCCGGGT TATGAGCGCA CACGCGACCT CCCGGACTAC	1853
25	ACTGGCATCC GTGACTCATT CGACCAGAAC ATGTCCACCG CTGTGGCCA CTTCGGCTCA	1913
	CTCTCCCACT CCTGCAGTAT CGTCACTAAG ACCTACCAGG GTTGGGAAGG CGTCACGAAC	1973
	GTCAACACGC CTTTCGGCCA ATTGCGCAC GCGGGCTCC TCAAGAATGA GGAGATCCTC	2033
30	TGCCTCGCG ACGACCTGGC CACCCGTCTC ACAGGTGTCT ACCCGGCCAC TGACAACCTC	2093
	GCGGCGCGG TTTCTGCCTT CGCCGCGAAC ATGCTGTCCT CCGTGTGAA GTCGGAGGCA	2153
35	ACGTCCTCCA TCATCAAGTC CGTTGGCGAG ACTGCCGTGC GCGCGGCTCA GTCCGGCTC	2213

GCGAAGCTAC CCGGACTGCT AATGAGTGTG CCAGGGAAGA TTGCCGCGCG TGTCCGCGCG 2273

GCGCGAGCGC GCCGCCGCGC CGCTCGTGCC AATTAGTTTG CTCGCTCCTG TTTCGCCGTT 2333

5 TCGTAAAACG GCGTGGTCCC GCACATTACG CGTACCCTAA AGACTCTGGT GAGTCCCCGT 2393

CGTTACACGA CGGGTCTGCC GCGGTTCGAT TCCATTCCCA AGCGGCAAGA AGGACGTAGT 2453

TAGCTCTGCG TCCCTCGGGA TACCA 2478

10

000001"659/450

5

(A) LENGTH: 157 amino acids

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

15

1 5 10 15

20 25 30

35 40 45

25

50 55 60

65 70 75 80

85 90 95

100 105 110

Ser Thr Ser Thr Pro Ile Pro Ser Ala Gly Thr Ser Ser Thr Leu Thr

115 120 125

Gln Arg Val Leu Gln Ser Leu Arg Ala Pro Ser Ala Ser Thr Arg Arg

130 135 140

5

Ser Leu Thr Ala Ser Ser Ser Ser Pro Ser Thr Gln Arg

145 150 155

10

(2) INFORMATION FOR SEQ ID NO:49:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 2478 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

15

(ii) MOLECULE TYPE: DNA (genomic)

20

(ix) FEATURE:

(A) NAME/KEY: CDS

(B) LOCATION: 366..2306

25

(x1) SEQUENCE DESCRIPTION: SEQ ID NO:49:

GTTTTCTTT CTTACCAAG TGTGGTAAAA TTAAACAAA GAAGAAAACC AGGACCGTAA 60

30

CCCGGCCCTT ACACACCTCG AGTCCGTGAC CACCGGATTA TAGTCGCCC ACCACACGGC 120

GCCTTTTCCG ACCACTCTCG AGAGTCGTTG GGAGTTTCGT CCGTGACCAC CCGGTTGGCA 180

GTCGACAGAC GCTTCGGGAC CACTAGAACC TCCTCGAGCG ACGCACACAC AGCACACACA 240

35

CCGCCTTAGC TGCACCTACG GCAGCGTTGA TAGCGCGGAT TTATGAGCGA GCACACCATC 300

20

5	AAT TGG GCC AAG GGA AAG ATC GAC CTC GAC TCC GAT TCC ATC GGC TGG	647
	Asn Trp Ala Lys Gly Lys Ile Asp Leu Asp Ser Asp Ser Ile Gly Trp	
	80 85 90	
10	TAC TTC AAG TAC CTT GAC CCA GCG GGT GCT ACA GAG TCT GCG CGC GCC	695
	Tyr Phe Lys Tyr Leu Asp Pro Ala Gly Ala Thr Glu Ser Ala Arg Ala	
	95 100 105 110	
15	GTC GGC GAG TAC TCG AAG ATC CCT GAC GGC CTC GTC AAG TTC TCC GTC	743
	Val Gly Glu Tyr Ser Lys Ile Pro Asp Gly Leu Val Lys Phe Ser Val	
	115 120 125	
20	GAC GCA GAG ATA AGA GAG ATC TAT AAC GAG GAG TGC CCC GTC GTC ACT	791
	Asp Ala Glu Ile Arg Glu Ile Tyr Asn Glu Glu Cys Pro Val Val Thr	
	130 135 140	
25	GAC GTG TCC GTC CCC CTC GAC GGC CGC CAG TGG AGC CTC TCG ATT TTC	839
	Asp Val Ser Val Pro Leu Asp Gly Arg Gln Trp Ser Leu Ser Ile Phe	
	145 150 155	
30	TCC TTT CCG ATG TTC AGA ACC GCC TAC GTC GCC GTA GCG AAC GTC GAG	887
	Ser Phe Pro Met Phe Arg Thr Ala Tyr Val Ala Val Ala Asn Val Glu	
	160 165 170	
35	AAC AAG GAG ATG TCG CTC GAC GTT GTC AAC GAC CTC ATC GAG TGG CTC	935
	Asn Lys Glu Met Ser Leu Asp Val Val Asn Asp Leu Ile Glu Trp Leu	
	175 180 185 190	
40	AAC AAT CTC GCC GAC TGG CGT TAT GTC GTT GAC TCT GAA CAG TGG ATT	983
	Asn Asn Leu Ala Asp Trp Arg Tyr Val Val Asp Ser Glu Gln Trp Ile	
	195 200 205	
45	AAC TTC ACC AAT GAC ACC ACG TAC TAC GTC CGC ATC CGC GTT CTA CGT	1031
	Asn Phe Thr Asn Asp Thr Thr Tyr Tyr Val Arg Ile Arg Val Leu Arg	
	210 215 220	

05677653 100300

	CCA ACC TAC GAC GTT CCA GAC CCC ACA GAG GGC CTT GTT CGC ACA GTC	1079
	Pro Thr Tyr Asp Val Pro Asp Pro Thr Glu Gly Leu Val Arg Thr Val	
	225 230 235	
5	TCA GAC TAC CGC CTC ACT TAT AAG GCG ATA ACA TGT GAA GCC AAC ATG	1127
	Ser Asp Tyr Arg Leu Thr Tyr Lys Ala Ile Thr Cys Glu Ala Asn Met	
	240 245 250	
	CCA ACA CTC GTC GAC CAA GGC TTT TGG ATC GGC GGC CAG TAC GCT CTC	1175
10	Pro Thr Leu Val Asp Gln Gly Phe Trp Ile Gly Gly Gln Tyr Ala Leu	
	255 260 265 270	
	ACC CCG ACT AGC CTA CCG CAG TAC GAC GTC AGC GAG GCC TAC GCT CTC	1223
15	Thr Pro Thr Ser Leu Pro Gln Tyr Asp Val Ser Glu Ala Tyr Ala Leu	
	275 280 285	
	CAC ACT TTG ACC TTC GCC AGA CCA TCC AGC GCC GCT GCA CTC GCG TTT	1271
	His Thr Leu Thr Phe Ala Arg Pro Ser Ser Ala Ala Ala Leu Ala Phe	
	290 295 300	
20	GTG TGG GCA GGT TTG CCA CAG GGT GGC ACT GCG CCT GCA GGC ACT CCA	1319
	Val Trp Ala Gly Leu Pro Gln Gly Gly Thr Ala Pro Ala Gly Thr Pro	
	305 310 315	
25	GCC TGG GAG CAG GCA TCC TCG GGT GGC TAC CTC ACC TGG CGC CAC AAC	1367
	Ala Trp Glu Gln Ala Ser Ser Gly Gly Tyr Leu Thr Trp Arg His Asn	
	320 325 330	
	GGT ACT ACT TTC CCA GCT GGC TCC GTT AGC TAC GTT CTC CCT GAG GGT	1415
30	Gly Thr Thr Phe Pro Ala Gly Ser Val Ser Tyr Val Leu Pro Glu Gly	
	335 340 345 350	
	TTC GCC CTT GAG CGC TAC GAC CCG AAC GAC GGC TCT TGG ACC GAC TTC	1463
35	Phe Ala Leu Glu Arg Tyr Asp Pro Asn Asp Gly Ser Trp Thr Asp Phe	
	355 360 365	

35

35

TCCCCGTCGT TACACGACGG GTCTGCCGCG GTTCGATTCC ATTCCCAAGC GGCAAGAAGG 2446

5

10

(A) LENGTH: 647 amino acids

(D) TOPOLOGY: linear

15

20

1 5 10 15

25

20 25 30

35 40 45

30

50 55 60

65 70 75 80

35

Ala Lys Gly Lys Ile Asp Leu Asp Ser Asp Ser Ile Gly Trp Tyr Phe

35

275 280 285
 Leu Thr Phe Ala Arg Pro Ser Ser Ala Ala Ala Leu Ala Phe Val Trp
 290 295 300
 5
 Ala Gly Leu Pro Gln Gly Gly Thr Ala Pro Ala Gly Thr Pro Ala Trp
 305 310 315 320
 Glu Gln Ala Ser Ser Gly Gly Tyr Leu Thr Trp Arg His Asn Gly Thr
 10 325 330 335
 Thr Phe Pro Ala Gly Ser Val Ser Tyr Val Leu Pro Glu Gly Phe Ala
 340 345 350
 15
 Leu Glu Arg Tyr Asp Pro Asn Asp Gly Ser Trp Thr Asp Phe Ala Ser
 355 360 365
 Ala Gly Asp Thr Val Thr Phe Arg Gln Val Ala Val Asp Glu Val Val
 370 375 380
 20
 Val Thr Asn Asn Pro Ala Gly Gly Gly Ser Ala Pro Thr Phe Thr Val
 385 390 395 400
 Arg Val Pro Pro Ser Asn Ala Tyr Thr Asn Thr Val Phe Arg Asn Thr
 25 405 410 415
 Leu Leu Glu Thr Arg Pro Ser Ser Arg Arg Leu Glu Leu Pro Met Pro
 420 425 430
 30
 Pro Ala Asp Phe Gly Gln Thr Val Ala Asn Asn Pro Lys Ile Glu Gln
 435 440 445
 Ser Leu Leu Lys Glu Thr Leu Gly Cys Tyr Leu Val His Ser Lys Met
 450 455 460
 35
 Arg Asn Pro Val Phe Gln Leu Thr Pro Ala Ser Ser Phe Gly Ala Val

000001"6694960

Ser Phe Asn Asn Pro Gly Tyr Glu Arg Thr Arg Asp Leu Pro Asp Tyr
485 490 495

Thr Gly Ile Arg Asp Ser Phe Asp Gln Asn Met Ser Thr Ala Val Ala
500 505 510

His Phe Arg Ser Leu Ser His Ser Cys Ser Ile Val Thr Lys Thr Tyr
515 520 525

Gln Gly Trp Glu Gly Val Thr Asn Val Asn Thr Pro Phe Gly Gln Phe
530 535 540

Ala His Ala Gly Leu Leu Lys Asn Glu Glu Ile Leu Cys Leu Ala Asp
545 550 555 560

Asp Leu Ala Thr Arg Leu Thr Gly Val Tyr Pro Ala Thr Asp Asn Phe
565 570 575

Ala Ala Ala Val Ser Ala Phe Ala Ala Asn Met Leu Ser Ser Val Leu
580 585 590

Lys Ser Glu Ala Thr Ser Ser Ile Ile Lys Ser Val Gly Glu Thr Ala
595 600 605

Val Gly Ala Ala Gln Ser Gly Leu Ala Lys Leu Pro Gly Leu Leu Met

610 615 620

Ser Val Pro Gly Lys Ile Ala Ala Arg Val Arg Ala Arg Arg Ala Arg
625 630 635 640

Arg Arg Ala Ala Arg Ala Asn
645

35

[illegible]

5

(D) TOPOLOGY: linear

15

20

25

30

5 10 15 20

35

ATA CCC CCT GAA CCT GAA GCA GGA TGG GAG ATG CTG GAG TGG CGT CAC 390

35

165 170 175 180

5 Trp Ser Leu Ser Ile Phe Ser Phe Pro Met Phe Arg Thr Ala Tyr Val

185 190 195

10 200 205 210

Ala Val Ala Asn Val Glu Asn Lys Glu Met Ser Leu Asp Val Val Asn

200 205 210

15

Asp Leu Ile Glu Trp Leu Asn Asn Leu Ala Asp Trp Arg Tyr Val Val

215 220 225

20 CGC ATC CGC GTT CTA CGT CCA ACC TAC GAC GTT CCA GAC CCC ACA GAG

Asp Ser Glu Gln Trp Ile Asn Phe Thr Asn Asp Thr Thr Tyr Tyr Val

230 235 240

CGC ATC CGC GTT CTA CGT CCA ACC TAC GAC GTT CCA GAC CCC ACA GAG 1062

Arg Ile Arg Val Leu Arg Pro Thr Tyr Asp Val Pro Asp Pro Thr Glu

245 250 255 260

	GGC CTT GTT CGC ACA GTC TCA GAC TAC CGC CTC ACT TAT AAG GCG ATA	1110
	Gly Leu Val Arg Thr Val Ser Asp Tyr Arg Leu Thr Tyr Lys Ala Ile	
	265 270 275	
5		
	ACA TGT GAA GCC AAC ATG CCA ACA CTC GTC GAC CAA GGC TTT TGG ATC	1158
	Thr Cys Glu Ala Asn Met Pro Thr Leu Val Asp Gln Gly Phe Trp Ile	
	280 285 290	
10		
	GGC GGC CAG TAC GCT CTC ACC CCG ACT AGC CTA CCG CAG TAC GAC GTC	1206
	Gly Gly Gln Tyr Ala Leu Thr Pro Thr Ser Leu Pro Gln Tyr Asp Val	
	295 300 305	
	AGC GAG GCC TAC GCT CTG CAC ACT TTG ACC TTC GCC AGA CCA TCC AGC	1254
15	Ser Glu Ala Tyr Ala Leu His Thr Leu Thr Phe Ala Arg Pro Ser Ser	
	310 315 320	
	GCC GCT GCA CTC GCG TTT GTG TGG GCA GGT TTG CCA CAG GGT GGC ACT	1302
	Ala Ala Ala Leu Ala Phe Val Trp Ala Gly Leu Pro Gln Gly Gly Thr	
20	325 330 335 340	
	GCG CCT GCA GGC ACT CCA GCC TGG GAG CAG GCA TCC TCG GGT GGC TAC	1350
	Ala Pro Ala Gly Thr Pro Ala Trp Glu Gln Ala Ser Ser Gly Gly Tyr	
	345 350 355	
25		
	CTC ACC TGG CGC CAC AAC GGT ACT ACT TTC CCA GCT GGC TCC GTT AGC	1398
	Leu Thr Trp Arg His Asn Gly Thr Thr Phe Pro Ala Gly Ser Val Ser	
	360 365 370	
30		
	TAC GTT CTC CCT GAG GGT TTC GCC CTT GAG CGC TAC GAC CCG AAC GAC	1446
	Tyr Val Leu Pro Glu Gly Phe Ala Leu Glu Arg Tyr Asp Pro Asn Asp	
	375 380 385	
	GGC TCT TGG ACC GAC TTC GCT TCC GCA GGA GAC ACC GTC ACT TTC CCG	1494
35	Gly Ser Trp Thr Asp Phe Ala Ser Ala Gly Asp Thr Val Thr Phe Arg	
	390 395 400	

	CAG GTC GCC GTC GAC GAG GTC GTT GTG ACC AAC AAC CCC GCC GGC GGC	1542
	Gln Val Ala Val Asp Glu Val Val Val Thr Asn Asn Pro Ala Gly Gly	
	405 410 415 420	
5	GGC AGC GCC CCC ACC TTC ACC GTG AGA GTG CCC CCT TCA AAC GCT TAC	1590
	Gly Ser Ala Pro Thr Phe Thr Val Arg Val Pro Pro Ser Asn Ala Tyr	
	425 430 435	
	ACC AAC ACC GTG TTT AGG AAC ACG CTC TTA GAG ACT CGA CCC TCC TCT	1638
10	Thr Asn Thr Val Phe Arg Asn Thr Leu Leu Glu Thr Arg Pro Ser Ser	
	440 445 450	
	CGT AGG CTC GAA CTC CCT ATG CCA CCT GCT GAC TTT GGA CAG ACG GTC	1686
	Arg Arg Leu Glu Leu Pro Met Pro Pro Ala Asp Phe Gly Gln Thr Val	
15	455 460 465	
	GCC AAC AAC CCG AAG ATC GAG CAG TCG CTT CTT AAA GAA ACA CTT GGC	1734
	Ala Asn Asn Pro Lys Ile Glu Gln Ser Leu Leu Lys Glu Thr Leu Gly	,
	470 475 480	
20	TGC TAT TTG GTC CAC TCC AAA ATG CGA AAC CCC GTT TTC CAG CTC ACG	1782
	Cys Tyr Leu Val His Ser Lys Met Arg Asn Pro Val Phe Gln Leu Thr	
	485 490 495 500	
25	CCA GCC AGC TCC TTT GGC GCC GTT TCC TTC AAC AAT CCG GGT TAT GAG	1830
	Pro Ala Ser Ser Phe Gly Ala Val Ser Phe Asn Asn Pro Gly Tyr Glu	
	505 510 515	
	CGC ACA CGC GAC CTC CCG GAC TAC ACT GGC ATC CGT GAC TCA TTC GAC	1878
30	Arg Thr Arg Asp Leu Pro Asp Tyr Thr Gly Ile Arg Asp Ser Phe Asp	
	520 525 530	
	CAG AAC ATG TCC ACC GCT GTG GCC CAC TTC CGC TCA CTC TCC CAC TCC	1926
	Gln Asn Met Ser Thr Ala Val Ala His Phe Arg Ser Leu Ser His Ser	
35	535 540 545	

	TGC AGT ATC GTC ACT AAG ACC TAC CAG GGT TGG GAA GGC GTC ACG AAC	1974
	Cys Ser Ile Val Thr Lys Thr Tyr Gln Gly Trp Glu Gly Val Thr Asn	
	550 555 560	
5	GTC AAC ACG CCT TTC GGC CAA TTC GCG CAC GCG GGC CTC CTC AAG AAT	2022
	Val Asn Thr Pro Phe Gly Gln Phe Ala His Ala Gly Leu Leu Lys Asn	
	565 570 575 580	
	GAG GAG ATC CTC TGC CTC GCC GAC GAC CTG GCC ACC CGT CTC ACA GGT	2070
10	Glu Glu Ile Leu Cys Leu Ala Asp Asp Leu Ala Thr Arg Leu Thr Gly	
	585 590 595	
	GTC TAC CCC GCC ACT GAC AAC TTC GCG GCC GCC GTT TCT GCC TTC GCC	2118
	Val Tyr Pro Ala Thr Asp Asn Phe Ala Ala Ala Val Ser Ala Phe Ala	
15	600 605 610	
	GCG AAC ATG CTG TCC TCC GTG CTG AAG TCG GAG GCA ACG TCC TCC ATC	2166
	Ala Asn Met Leu Ser Ser Val Leu Lys Ser Glu Ala Thr Ser Ser Ile	
	615 620 625	
20	ATC AAG TCC GTT GGC GAG ACT GCC GTC GGC GCG GCT CAG TCC GGC CTC	2214
	Ile Lys Ser Val Gly Glu Thr Ala Val Gly Ala Ala Gln Ser Gly Leu	
	630 635 640	
25	GCG AAG CTA CCC GGA CTG CTA ATG AGT GTA CCA GGG AAG ATT GCC GCG	2262
	Ala Lys Leu Pro Gly Leu Leu Met Ser Val Pro Gly Lys Ile Ala Ala	
	645 650 655 660	
	CGT GTC CGC GCG CGC CGA GCG CGC CGC CGC GCC GCT CGT GCC AAT	2307
30	Arg Val Arg Ala Arg Arg Ala Arg Arg Arg Ala Ala Arg Ala Asn	
	665 670 675	
	TAGTTTGCTC GCTCCTGTTT CGCCGTTTCG TAAAACGGCG TGGTCCCGCA CATTACGCGT	2367
35	ACCCTAAAGA CTCTGGTGAG TCCCCGTCGT TACACGACGG GTCTGCCGCG GTTCGATTCC	2427

2479

5

10

(B) TYPE: amino acid

(ii) MOLECULE TYPE: protein

15

1 5 10 15

20

20 25 30

35 40 45

25

50 55 60

30

65 70 75 80

85 90 95

35

100 105 110

35

5

10

15

20

25

30

35

Glu Thr Leu Gly Cys Tyr Leu Val His Ser Lys Met Arg Asn Pro Val
485 490 495

Phe Gln Leu Thr Pro Ala Ser Ser Phe Gly Ala Val Ser Phe Asn Asn
 500 505 510

5 Pro Gly Tyr Glu Arg Thr Arg Asp Leu Pro Asp Tyr Thr Gly Ile Arg
 515 520 525

Asp Ser Phe Asp Gln Asn Met Ser Thr Ala Val Ala His Phe Arg Ser
 530 535 540

10 Leu Ser His Ser Cys Ser Ile Val Thr Lys Thr Tyr Gln Gly Trp Glu
 545 550 555 560

Gly Val Thr Asn Val Asn Thr Pro Phe Gly Gln Phe Ala His Ala Gly
 565 570 575

15 Leu Leu Lys Asn Glu Glu Ile Leu Cys Leu Ala Asp Asp Leu Ala Thr
 580 585 590

Arg Leu Thr Gly Val Tyr Pro Ala Thr Asp Asn Phe Ala Ala Ala Val
 20 595 600 605

Ser Ala Phe Ala Ala Asn Met Leu Ser Ser Val Leu Lys Ser Glu Ala
 610 615 620

25 Thr Ser Ser Ile Ile Lys Ser Val Gly Glu Thr Ala Val Gly Ala Ala
 625 630 635 640

Gln Ser Gly Leu Ala Lys Leu Pro Gly Leu Leu Met Ser Val Pro Gly
 645 650 655

30 Lys Ile Ala Ala Arg Val Arg Ala Arg Arg Ala Arg Arg Arg Ala Ala
 660 665 670

Arg Ala Asn
 35 675

5

(A) LENGTH: 59 base pairs

(C) STRANDEDNESS: single

10

(X1) SEQUENCE DESCRIPTION: SEQ ID NO:53:

15

GGGGATCCAC AGTTCTGCCT CCCCCGGACG GTAAATATAG GGGAACCATG GTCTAGAGG

59

CLAIMS

1. A capsovector for use in controlling insect pests, the capsovector comprising a capsid protein of an insect small RNA virus encapsidating an insecticidal protein toxin,
5 the capsid protein protecting the protein toxin from inactivation in the gut of an insect following ingestion of the capsovector by the insect.
2. A capsovector as claimed in claim 1 in which the insect small RNA virus is HaSV.
- 10 3. A capsovector as claimed in claim 1 in which the capsid protein is P71 (SEQ ID No. 50)
4. A capsovector as claimed in claim 1 in which the insecticidal toxin is of plant
15 origin.
5. A capsovector as claimed in claim 1 in which the insecticidal toxin is Ricin A or diptheria toxin.
- 20 6. A capsovector for use in controlling insect pests, the capsovector comprising a capsid protein of an insect small RNA virus encapsidating a nucleic acid molecule which is insecticidal or which encodes an insecticidal protein toxin, the capsid protein protecting the protein toxin from inactivation in the gut of an insect following
25 ingestion of the capsovector by the insect.
7. A capsovector as claimed in claim 6 in which the nucleic acid is RNA.
8. A capsovector as claimed in claim 7 in which the insect small RNA virus is HaSV.

30

0067653-100300

9. A capsovector as claimed in claim 7 in which the capsid protein is P71 (SEQ ID No. 50)
10. A capsovector as claimed in claim 7 in which the insecticidal toxin is of plant origin.
11. A capsovector as claimed in claim 7 in which the insecticidal toxin is Ricin A or diphtheria toxin.
12. A capsovector as claimed in claim 7 in which the RNA is an antisense sequence, a ribozyme or a mimicking structure.
13. A capsovector as claimed in claim 12 in which the mimicking structure is RNA hybridised so as to at least partially form double stranded RNA.
14. A capsovector as claimed in claim 6 in which the capsovector includes a further nucleic acid sequence, the further nucleic acid sequence encoding the capsid protein.
15. An isolated nucleic acid molecule comprising a first sequence encoding at least one capsid protein of an insect small RNA virus and a second sequence which is insecticidal or which encodes an insecticidal protein toxin.
16. An isolated nucleic acid molecule as claimed in claim 15 in which the nucleic acid is RNA.
17. An isolated nucleic acid molecule as claimed in claim 15 in which the insect small RNA virus is HaSV.
18. An isolated nucleic acid molecule as claimed in claim 15 in which the capsid protein is P71 (SEQ ID No. 50)

19. An isolated nucleic acid molecule as claimed in claim 15 in which the insecticidal toxin is of plant origin.

20. An isolated nucleic acid molecule as claimed in claim 15 in which the insecticidal toxin is Ricin A.

21. An isolated nucleic acid molecule as claimed in claim 15 in which the second sequence is an antisense sequence, a ribozyme or a mimicking structure.

22. An isolated nucleic acid molecule as claimed in claim 21 in which the mimicking structure is double stranded RNA.

23. An isolated nucleic acid molecule as claimed in claim 15 in which the insecticidal toxin is less toxic to plants than insects.

24. A transgenic plant resistant to insect attack comprising a genome or subgenome capable of expressing the nucleic acid molecule as claimed in claim 15 such that the transgenic plant produces capsid protein in which is encapsidated the nucleic acid molecule.

25. An isolated nucleic acid molecule comprising a first sequence encoding at least one capsid protein of an insect small RNA virus, a second sequence which is insecticidal or which encodes an insecticidal protein toxin and a third sequence positioned between the first and second sequence, the third sequence directing expression of the second sequence in an insect pest.

26. An isolated nucleic acid molecule as claimed in claim 25 in which the third sequence is an IRES.

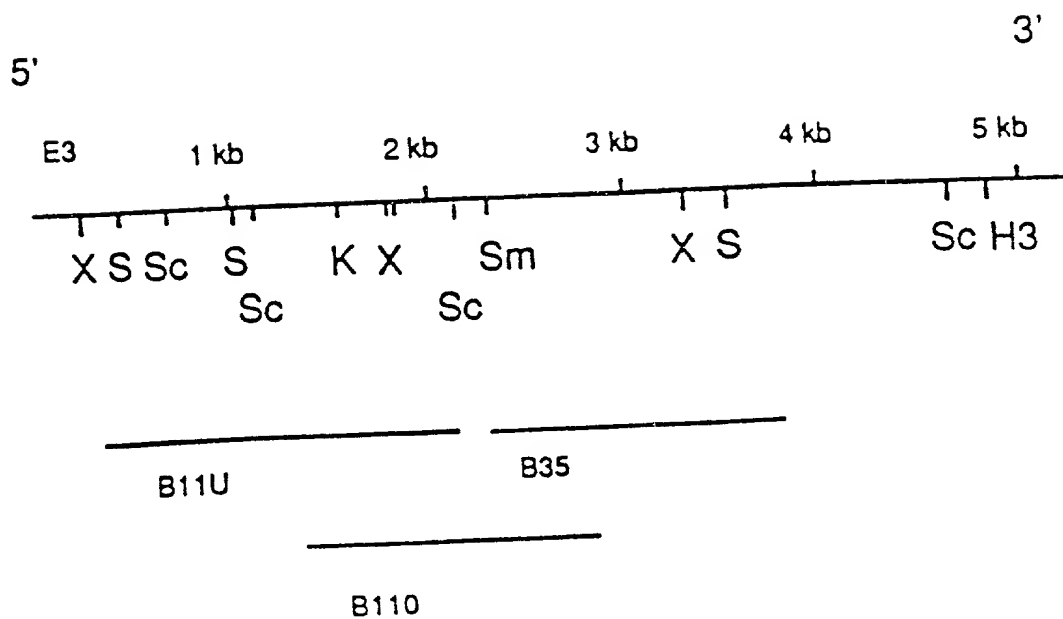
27. An isolated nucleic acid molecule as claimed in claim 25 in which the nucleic acid is RNA.
28. An isolated nucleic acid molecule as claimed in claim 25 in which the insect
5 small RNA virus is HaSV.
29. An isolated nucleic acid molecule as claimed in claim 25 in which the capsid protein is P71 (SEQ ID No. 50)
- 10 30. An isolated nucleic acid molecule as claimed in claim 25 in which the insecticidal toxin is of plant origin.
31. An isolated nucleic acid molecule as claimed in claim 25 in which the insecticidal toxin is Ricin A or diptheria toxin.
15
32. A transgenic plant resistant to insect attack comprising a genome or subgenome capable of expressing the nucleic acid molecule as claimed in claim 25 such that the transgenic plant produces capsid protein in which is encapsidated the nucleic acid molecule.

20

	1980	1981	1982	1983	1984	1985	1986	1987	1988	1989	1990	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021	2022	2023	2024	2025	2026	2027	2028	2029	2030	2031	2032	2033	2034	2035	2036	2037	2038	2039	2040	2041	2042	2043	2044	2045	2046	2047	2048	2049	2050	2051	2052	2053	2054	2055	2056	2057	2058	2059	2060	2061	2062	2063	2064	2065	2066	2067	2068	2069	2070	2071	2072	2073	2074	2075	2076	2077	2078	2079	2080	2081	2082	2083	2084	2085	2086	2087	2088	2089	2090	2091	2092	2093	2094	2095	2096	2097	2098	2099	2100	2101	2102	2103	2104	2105	2106	2107	2108	2109	2110	2111	2112	2113	2114	2115	2116	2117	2118	2119	2120	2121	2122	2123	2124	2125	2126	2127	2128	2129	2130	2131	2132	2133	2134	2135	2136	2137	2138	2139	2140	2141	2142	2143	2144	2145	2146	2147	2148	2149	2150	2151	2152	2153	2154	2155	2156	2157	2158	2159	2160	2161	2162	2163	2164	2165	2166	2167	2168	2169	2170	2171	2172	2173	2174	2175	2176	2177	2178	2179	2180	2181	2182	2183	2184	2185	2186	2187	2188	2189	2190	2191	2192	2193	2194	2195	2196	2197	2198	2199	2200	2201	2202	2203	2204	2205	2206	2207	2208	2209	2210	2211	2212	2213	2214	2215	2216	2217	2218	2219	2220	2221	2222	2223	2224	2225	2226	2227	2228	2229	2230	2231	2232	2233	2234	2235	2236	2237	2238	2239	2240	2241	2242	2243	2244	2245	2246	2247	2248	2249	2250	2251	2252	2253	2254	2255	2256	2257	2258	2259	2260	2261	2262	2263	2264	2265	2266	2267	2268	2269	2270	2271	2272	2273	2274	2275	2276	2277	2278	2279	2280	2281	2282	2283	2284	2285	2286	2287	2288	2289	2290	2291	2292	2293	2294	2295	2296	2297	2298	2299	2300	2301	2302	2303	2304	2305	2306	2307	2308	2309	2310	2311	2312	2313	2314	2315	2316	2317	2318	2319	2320	2321	2322	2323	2324	2325	2326	2327	2328	2329	2330	2331	2332	2333	2334	2335	2336	2337	2338	2339	2340	2341	2342	2343	2344	2345	2346	2347	2348	2349	2350	2351	2352	2353	2354	2355	2356	2357	2358	2359	2360	2361	2362	2363	2364	2365	2366	2367	2368	2369	2370	2371	2372	2373	2374	2375	2376	2377	2378	2379	2380	2381	2382	2383	2384	2385	2386	2387	2388	2389	2390	2391	2392	2393	2394	2395	2396	2397	2398	2399	2400	2401	2402	2403	2404	2405	2406	2407	2408	2409	2410	2411	2412	2413	2414	2415	2416	2417	2418	2419	2420	2421	2422	2423	2424	2425	2426	2427	2428	2429	2430	2431	2432	2
--	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	---

5

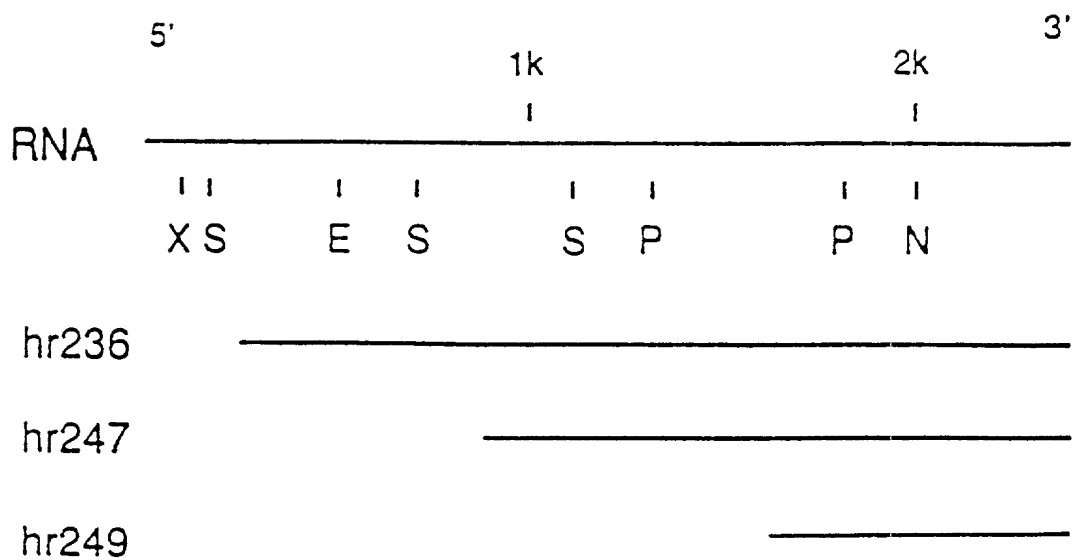
Map of HaSV RNA 1 clones



H3=Hind3, K=Kpn1, Sc=Sac1, S=Sal1, Sm=Sma1, X=Xho1

III. 1a.

Map of HaSV RNA 2 clones



E=EcoR1, N=Not1, P=Pst1, S=Sal1. X=Xho1

III. 1b.

```

10          30          50
GTTCTGCCCTCCCGGACGGTAATATAGGGGAACAATGTACGCCGAAGCGACAGACGTG
-----+-----+-----+-----+-----+-----+-----+
M Y A K A T D V
replicase start
70          90          110
GCGCGTGTCTACGCCCGGCAGATGTCCCTACGCCGAACGTACTGCAGCAGAGAGCAGTC
-----+-----+-----+-----+-----+-----+-----+
A R V Y A A A D V A Y A N V L Q Q R A V

130          150          170
AAGTTGGACTTCGCCCGCCACTGAAGGCACCTAGAAACCCCTCCACAGACTGTACTATCCG
-----+-----+-----+-----+-----+-----+-----+
K L D F A A P P L K A L E T L H R L Y Y P

190          210          230
CTGCGCTTCAAAGGGGCACCTTTACCCCGACACAAACCCGATCCTGGCCGGCACCAA
-----+-----+-----+-----+-----+-----+-----+
L R F K G G T L P P T Q H P I L A G H Q

250          270          290
CGTGTCCGAGAAGAGGTTCTGCACAATTTCGCCAGGGGACGTAGCACAGTCTCGAGATA
-----+-----+-----+-----+-----+-----+-----+
R V A E E V L H N F A R G R S T V L E I

```

III. 2.1.

310 330 350
GGGCCGTCTGTGCACAGCGCACTTAAGCTACATGGGGCACCGAACGCCCCCGTCGCAGAC
-----+-----+-----+-----+-----+
G P S L H S A L K L H G A P N A P V A D

370 390 410
TATCACGGGTGCACCAAGTACGGCACCCCGGACGGCTCGGACACATACGGCCTTAGAG
-----+-----+-----+-----+-----+
Y H G C T K Y G T R D G S R H I T A L E

430 450 470
TCTAGATCCGTGCGCACAGCGCGGCCGAGTTCAAGGCCGACGCCCTCACTGCTCGCCAAC
-----+-----+-----+-----+-----+
S R S V A T G R P E F K A D A S L L A N

490 510 530
GGCATTGCCCTCCCGCACCTTCTGCGTCGACGGAGTCGGCTCTTGGCGGTTCAAATCGCGC
-----+-----+-----+-----+-----+
G I A S R T F C V D G V G S C A F K S R

550 570 590
GTTGGAATTGCCAATCACTCCCTCTATGACGTGACCCCTAGAGGAGCTGGCCAATGCGTTT
-----+-----+-----+-----+-----+
V G I A N H S L Y D V T L E E L A N A F

III. 2.2

610 630 650
GAGAACCGGACTTCACATGGTCCGCCGTTTCATGCACATGCCAGAAGAGCTGCTCTAC
-----+-----+-----+-----+-----+
E N H G L H M V R A F M H M P E E L L Y

670 690 710
ATGGACAACGTGGTTAATGCCGAGCTCGGCTACCGCTTCCACGTTATTGAAGAGCCCTATG
-----+-----+-----+-----+-----+
M D N V V N A E L G Y R F H V I E E P M

730 750 770
GCTGTGAAGGACTGCCGATTCAGGGGGGACCTCCGTCCTCCACTTCCCTGAGTTGGAC
-----+-----+-----+-----+-----+
A V K D C A F Q G G D L R L H F P E L D

790 810 830
TTCATCAACGAGAGCCAAGAGCGGCGCATCGAGAGGCTGGCCCGCCGGCTCCTACTCC
-----+-----+-----+-----+-----+
F I N E S Q E R R I E R L A A R G S Y S

850 870 890
AGACGGCCGTCATTTTCTCCGGCAGCAGACTGGGGTGATGCGTACTTACAGACTTC
-----+-----+-----+-----+-----+
R R A V I F S G D D D W G D A Y L H D F

III 23

910 930 950
CACACATGGCTCGCCTACCTACTGGTGAGGAACCTACCCACTCCGTTTGGTTTCTCACTC
-----+-----+-----+-----+-----+-----+
H T W L A Y L L V R N Y P T P F G F S L

970 990 1010
CATATAGAAGTCCAGAGGCGCCACGGCTCCAGCATTGAGCTGCGCATCACTGCGCGGCCA
-----+-----+-----+-----+-----+-----+
H I E V Q R R H G S S I E L R I T R A P

1030 1050 1070
CCTGGAGACCGCATGCTGGCCGTCGTCCCAAGGACGTCCCAAGGCCTCTGCAGAAATCCCA
-----+-----+-----+-----+-----+-----+
P G D R M L A V V P R T S Q G L C R I P

1090 1110 1130
AACATCTTTTATTACGCCGACGCGTCGGGCACTGAGCATAAGACCATCCTTACGTCACAG
-----+-----+-----+-----+-----+-----+
N I F Y Y A D A S G T E H K T I L T S Q

1150 1170 1190
CACAAAGTCAACATGCTGCTCAATTTTATGCAAAACGCGTCCTGAGAAGGAAGTACGAC
-----+-----+-----+-----+-----+-----+
H K V N M L L N F M Q T R P E K E L V D

1210 1230 1250
 ATGACCGTCTTGATGTCGTCGGCGGCTAGGCTGCGCGGATCGTGGTCGCCCTCAGAA
 -----+-----+-----+-----+-----+-----+-----+
 M T V L M S F A R A R L R A I V V A S E

1270 1290 1310
 GTCACCGAGAGCTCCTGGAAACATCTACCGGCTGACCTGGTCCGCACTGTCGTCTCTT
 -----+-----+-----+-----+-----+-----+-----+
 V T E S S W N I S P A D L V R T V V S L

1330 1350 1370
 TACGTCTCCACATCATCGAGCGCCGAAGGCTGCGTGTCTCAAGACCGCCAAGGAC
 -----+-----+-----+-----+-----+-----+-----+
 Y V L H I I E R R R A A V A V K T A K D

1390 1410 1430
 GACGTCTTTGGAGAGACTTCGTTCTGGGAGAGTCTCAAGCACGTCCTTGGGCTCCTGTTGC
 -----+-----+-----+-----+-----+-----+-----+
 D V F G E T S F W E S L K H V L G S C C

1450 1470 1490
 GGTCTGCGCAACCTCAAGGCACCGACGTCGTCCTTACTAAGCGCGTCGTCGATAAGTAC
 -----+-----+-----+-----+-----+-----+-----+
 G L R N L K G T D V V F T K R V V D K Y

1510 1530 1550
CGAGTCCACTCGGAGACATAATCTGCGACGTCCGCCCTGTCCCCCTGAACAGGTCGGC
-----+-----+-----+-----+-----+-----+
R V H S L G D I I C D V R L S P E Q V G

1570 1590 1610
TTCCCTGCCGTCGCGGTACCACTGCCCGCGTCTTTCACGACAGGGAAGCTTGAGGTC
-----+-----+-----+-----+-----+-----+
F L P S R V P P A R V F H D R E E L E V

1630 1650 1670
CTTCGCGAAGCTGGCTGTACAACGAACGTCCGGTACCTTCCACTCCTCCTGTGGAGGAG
-----+-----+-----+-----+-----+-----+
L R E A G C Y N E R P V P S T P P V E E

1690 1710 1730
CCCCAAGGTTTCGACGCCGACTTGTGGCACGCGGACCGGGCCTCACTCCCCGAGTACCGC
-----+-----+-----+-----+-----+-----+
P Q G F D A D L W H A T A A S L P E Y R

1750 1770 1790
GCCACCTTGACGAGGTCTCAACACCGACGTCAAGCAGCTCAAGATCACCCCTCGAGAAC
-----+-----+-----+-----+-----+-----+
A T L Q A G L N T D V K Q L K I T L E N

1810 1830 1850
 GCCCTCAAGACCATCGACGGGCTCACCCCTCTCCCCAGTCAGAGGCCCTCGAGATGTACGAG
 -----+-----+-----+-----+-----+-----+-----+
 A L K T I D G L T L S P V R G L E M Y E

1870 1890 1910
 GGCCCGCCAGGCGGCAAGACGGGCACCCCTCATCGCCGCCCTTGAGGCCGCGGGCGGT
 -----+-----+-----+-----+-----+-----+-----+
 G P P G S G K T G T L I A A L E A A G G

1930 1950 1970
 AAAGCACTTACGTGGCACCCACAGAGAACTGAGAGAGGCTATGGACCGCGGATCAAA
 -----+-----+-----+-----+-----+-----+-----+
 K A L Y V A P T R E L R E A M D R R I K

1990 2010 2030
 CCGCCGTCGGCTACGCAACATGTGCGCCCTTGGGATTCTCCGTGCGTGCACCGCC
 -----+-----+-----+-----+-----+-----+-----+
 P P S A S A T Q H V A L A I L R R A T A

2050 2070 2090
 GAGGGCGCCCTTTCGCTACCGTGGTTATCGACGAGTGCTTCATGTCCCGCTCGTGAC
 -----+-----+-----+-----+-----+-----+-----+
 E G A P F A T V V I D E C F M F P L V Y

2110 2130 2150
 GTCGCCATCGTGCACGCCCTTGTCCCCGAGCTCACGAATAGTCCTTGTAGGGGACGTCCAC
 -----+-----+-----+-----+-----+-----+-----+
 V A I V H A L S P S S R I V L V G D V H

2170 2190 2210
 CAAATCGGGTTATAGACTTCCAAGGCACAAAGCGCGAACAATGCCGCTCGTTCCGCGACGTC
 -----+-----+-----+-----+-----+-----+-----+
 Q I G F I D F Q G T S A N M P L V R D V

2230 2250 2270
 GTTAAGCAGTGCCGTCGGCGCACTTCAACCAACCAAGCGCTGTCCGGCCGACGTCGTT
 -----+-----+-----+-----+-----+-----+-----+
 V K Q C R R R T F N Q T K R C P A D V V

2290 2310 2330
 GCCACCACGTTTCCAGAGCTTGTACCCCGGGTGCACAACCACTCAGGGTGCGTCGCA
 -----+-----+-----+-----+-----+-----+-----+
 A T T F F Q S L Y P G C T T T S G C V A

2350 2370 2390
 TCCATCAGCCACGTGCCCCAGACTACCGCAACAGCCAGCGCAACGCTCTGCTTCACG
 -----+-----+-----+-----+-----+-----+-----+
 S I S H V A P D Y R N S Q A Q T L C F T

2410 2430 2450
CAGGAGGAAAAGTCGGCCACGGGGCTGAGGGCGCGATGACTGTGCACGAAGCGCAGGGA
-----+-----+-----+-----+-----+-----+
Q E E K S R H G A E G A M T V H E A Q G

2470 2490 2510
CGCACTTTTGGCTGTCTGCTATCTGCATTACAACGGCTCCACAGCAGAGCAGAAGCTCCTC
-----+-----+-----+-----+-----+-----+
R T F A S V I L H Y N G S T A E Q K L L

2530 2550 2570
GCTGAGAAGTCGCACCTTCTAGTCGGCATCACGGCCACACCAACCACCTGTACATCCGC
-----+-----+-----+-----+-----+-----+
A E K S H L L V G I T R H T N H L Y I R

2590 2610 2630
GACCCGACAGGTGACATTGAGAGACAACCTCAACCATAGCGCGAAAGCCGAGGTGTTACA
-----+-----+-----+-----+-----+-----+
D P T G D I E R Q L N H S A K A E V F T

2650 2670 2690
GACATCCCTGCACCCCTGGAGATCACGACTGTCAAACCGAGTGAAGAGTGCAGCGCAAC
-----+-----+-----+-----+-----+-----+
D I P A P L E I T T V K P S E E V Q R N

3010 3030 3050
CTACACGAAGCTAAGGAGGACGTCAAACGCAATGCTAAACTCGCTTGACCCGACATGGGAC
-----+-----+-----+-----+-----+-----+
L H E A K E D V K R M L N S L D R H W D

3070 3090 3110
TGGACTGTCACTGAAGACGCCCGTGACCCGAGCTGTCTTCGAGACCCAGCTCAAGTTCACC
-----+-----+-----+-----+-----+-----+
W T V T E D A R D R A V F E T Q L K F T

3130 3150 3170
CAACGGCGGCACCGTCGAAGACCTGCTGGAGCCAGACGCCCTACATCCGTGACATA
-----+-----+-----+-----+-----+-----+
Q R G G T V E D L L E P D D P Y I R D I

3190 3210 3230
GACTTCCTTATGAAGACTCAGCAGAAAGTGTGCGCCCAAGCCGATCAATACGGGCAAGGTC
-----+-----+-----+-----+-----+-----+
D F L M K T Q Q K V S P K P I N T G K V

3250 3270 3290
GGGCAGGGGATCGCCGCTCACTCAAAGTCTCTCAACTTCGTCTCGCCGCTTGATACGC
-----+-----+-----+-----+-----+-----+
G Q G I A A H S K S L N F V L A A W I R

III. 2.11

3310 3330 3350
 ATACTCGAGAGATACTCCGTACCGGAGCCGACGGTCCGGTACAGCAACGGTCTCCCC
 -----+-----+-----+-----+-----+-----+
 I L E E I L R T G S R T V R Y S N G L P

3370 3390 3410
 GACGAAGAAGGCCATGCTGCTCGAAGCGAAGATCAATCAAGTCCACACGCCACGTTC
 -----+-----+-----+-----+-----+-----+
 D E E E A M L L E A K I N Q V P H A T F

3430 3450 3470
 GTCTCGCGGACTGGACCGAGTTTGACACCGCCACAAATAACACGAGTGAGCTGCTCTTC
 -----+-----+-----+-----+-----+-----+
 V S A D W T E F D T A H N N T S E L L F

3490 3510 3530
 GCCGCCCTTTAGAGCGCATCGGCACGCCCTGCAGCTGCCGTTAATCTATTCAGAGAACGG
 -----+-----+-----+-----+-----+-----+
 A A L L E R I G T P A A A V N L F R E R

3550 3570 3590
 TGTGGAAACGCACCTTGCGAGCGAAGGCTCTAGGCTCCGTTGAAGTCGACGGTCTGCTC
 -----+-----+-----+-----+-----+-----+
 C G K R T L R A K G L G S V E V D G L L

3910 3930 3950
GAACTCCTTACTCAAGTACGTGGAGGCTGTGAGAGACATCACCAAGGCTGGAGTGAC
-----+-----+-----+-----+-----+-----+
E L L Y S K Y V E A V R D I T K G W S D

3970 3990 4010
GCCCCTACACAGCCTCCTGTGCCACATGTCAGCATGCTACTACAATTACGCCGCCGGAG
-----+-----+-----+-----+-----+-----+
A R Y H S L L C H M S A C Y Y N Y A P E

4030 4050 4070
TCTGCCGCGTACATCGACGCTGTTGCTTGGGCGCGGCGACTTCCCGTTTGAA
-----+-----+-----+-----+-----+-----+
S A A Y I I D A V V R F G R G D F P F E

4090 4110 4130
CAACTGCGGTGGTGCGTCCCATGTGCAGGACCCGACGCTTACAGCAGCACGTATCCG
-----+-----+-----+-----+-----+-----+
Q L R V V R A H V Q A P D A Y S S T Y P

4150 4170 4190
GCTAACGTGCGCGCATCGTGCCCTTGACCACGCTCTCGAGCCCCGCCAGGCCGCCCGCCG
-----+-----+-----+-----+-----+-----+
A N V R A S C L D H V F E P R Q A A A P

III. 2.14

4210 4230 4250
GCAGGTTTCGTTGCGACATGTCCGAAGCCGGAAACGCCCTTCTTCACTTACCGCGAAAGCT
-----+-----+-----+-----+-----+-----+
A G F V A T C A K P E T P S S L T A K A
M C E A G N A F F T Y R E S W
P11a start
4270 4290 4310
GGTGTTCGCGACTACAAGCCACGTTGCGACTGGGACTGCGCCCCCGGAGTCTCCATGG
-----+-----+-----+-----+-----+-----+
G V S A T T S H V A T G T A P P E S P W
C F C D Y K P R C D W D C A P G V S M G
4330 4350 4370
GATGCACCTGCAGCCAAACAGCTTTTCGGAGTTATTGACACCGGAGACCCCGTCCACATCA
-----+-----+-----+-----+-----+-----+
D A P A A N S F S E L L T P E T P S T S
C T C S Q Q L F G V I D T G D P V H I I
4390 4410 4430
TCCTCGCCGTCATCGTCTTCATCGGACTCCCTCTACATCGTGTGAAGGTCGCTCAGTGGT
-----+-----+-----+-----+-----+-----+
S S P S S S S D S S T S C G R S L S G
L A V I V F I G L L Y I V W K V A Q W W

III 2.15

4450 4470 4490
 GGAGACCCGCAAGGACCACAGAAGACTTGAACAGCAGAAAGCCGCTTCGCAAGACAGG
 -----+-----+-----+-----+-----+-----+
 G D T A R T T E D L N S R K P P S Q D R
 R H R K D H R R L E Q Q K A A F A R Q A

4510 4530 4550
 CAATCACGCTCGTCTGAATGTCTGGACAGAAGCGGAGAAAGACAGGCAGTTCGTTAACT
 -----+-----+-----+-----+-----+-----+
 Q S R S S E C L D R S G E R T G S S L T
 I T L V * M S G Q K R R K D R Q F V N C
 P11b start

4570 4590 4610
 GCCCCACTGCTCCGAGCCCCCTCATTTCTCATTTTCGGAAAGAGCTCGACTGGCGACCGGG
 -----+-----+-----+-----+-----+-----+
 A P T A P S P S F S F S E R A R L A T G
 P H C S E P L I L I F G K S S T G D R A

4630 4650 4670
 CCGACTGTCGCCGCTGGACATCACCTTCGGCAACCCCATCCTGGCCACGACGAGGTT
 -----+-----+-----+-----+-----+-----+
 P T V A A A T S P S A T P S C A T D Q V
 D C R R C D I T F G N P I L R H G P G C

4690 4710 4730
GCCGCCGAGCACGCCGACTTTGCGCCTTCTGGTTCAGTCTGCCCGTCTGTC
-----+-----+-----+-----+-----+-----+
A A R T T P D F A P F L G S Q S A R A V
R E D H A G L C A F P G F P V C P C C L

4750 4770 4790
TCGAAGCCGTACCGGCCCCACGACTGCCCGTTGGAAGAAGTCACCCCGTCCACGCG
-----+-----+-----+-----+-----+-----+
S K P Y R P P T T A R W K E V T P L H A
E A V P A P H D C P L E R S H P A P R V

4810 4830 4850
TGGAAGGCGTGACCGGAGACCGACCGGAAGTCAGGGAGGACCCGGAGACGCGGTC
-----+-----+-----+-----+-----+-----+
W K G V T G D R P E V R E D P E T A A V
E G R D R R P T G S Q G G P G D S G G R

4870 4890 4910
GTCCAGGCTCTGATCAGCGCGCGTTATCTCAGAAAGCTTCTCCGACGATCC
-----+-----+-----+-----+-----+-----+
V Q A L I S G R Y P Q K T K L S S D A S
P G S D Q R P L S S E D E A F L R R I Q

4930 4950 4970
AAAGGCTACTCAAGAACTAAGGATGCTCACAATCCACCTCTTTCTCTGCCCGAGTGCG
-----+-----+-----+-----+-----+-----+-----+
K G Y S R T K G C S Q S T S F P A P S A
R L L K N * M L T I H L F S C P E C G
P14 start
4990 5010 5030
GATTACGAGGCCGCGACTGCCAGACAGTCCGAGTCTGCCGCGCTGCAGAGATGGCG
-----+-----+-----+-----+-----+-----+-----+
D Y Q A R D C Q T V R V C R A A A E M A
L P G P R L P D S P S L P R R C R D G A
5050 5070 5090
CGCTCATGTATTCACGAGCCGTGGCTTCATCTGCCGCCAGTGCCGACTTGAAGCGCAT
-----+-----+-----+-----+-----+-----+-----+
R S C I H E P L A S S A A S A D L K R I
L M Y S R A V G F I C R Q C R L E A H T
5110 5130 5150
CGCTCTACCTCGGACTCTGTCCCGATGTAAGATCAGCAAGAGCGCATGAAGGAACAA
-----+-----+-----+-----+-----+-----+-----+
R S T S D S V P D V K I S K S A *
L Y L G L C S R C K D Q Q E R M K E Q N

五-2.18

21/56

*

5230 5250 5270
TCCCTCAACGTTACTCGTTGAGTCTGCTGCGGTTCCGATTCCATTCCCAAGCAGCAAGGGT
-----+-----+-----+-----+-----+-----+-----+-----+

5290 5310
GGGCAACTAGTACGGCGCCCTGGGATACCA
-----+-----+-----+-----

2.19


```

370          390          410
GCAGGATGGAGATGCTGGAGTGGCGTCAACAGCGACCTCACAAACCGTCGCGGAACCCGTA
-----+-----+-----+-----+-----+-----+-----+
A G W E M L E W R H S D L T T V A E P V
M G D A G V A S Q R P H N R R G T R N
P71 start
430          450          470
ACGTCGGGTCAGCGCCAACACCGTCACCGTCAATGGTAGAAGAAACCAACGGCGTCGGA
-----+-----+-----+-----+-----+-----+-----+
T F G S A P T P S P S M V E E T N G V G
V R V S A N T V T V N G R R N Q R R R T

490          510          530
CCGGAAGGCAAGTTTCTCCCCCTGACAATTTCACCGCTGCTGCACAAGACCTCGCGCAA
-----+-----+-----+-----+-----+-----+-----+
P E G K F L P L T I S P L L H K T S R K
G R Q V S P P D N F T A A A Q D L A Q S

550          570          590
GCCTTGACGCCCAACACCGTCACCTTCCCGCTAACATCTCTAGCATGCCCGAATTCCGGA
-----+-----+-----+-----+-----+-----+-----+
A L T P T P S L S P L T S L A C P N S G
L D A N T V T F P A N I S S M P E F R N

```

III.3a.2

610 630 650
 ATTGGCCAGGGAAGATCGACCTCGACTCCGATTCGCTGGCTGGTACTTCAAGTACC
 -----+-----+-----+-----+-----+-----+-----+
 I G P R E R S T S T P I P S A G T S S T
 W A K G K I D L D S D S I G W Y F K Y L

670 690 710
 TTGACCCAGCGGTGCTACAGAGTCTGCGCGCGCGCTCGGCGAGTACTCGAAGATCCCTG
 -----+-----+-----+-----+-----+-----+-----+
 L T Q R V L Q S L R A P S A S T R R S L
 D P A G A T E S A R A V G E Y S K I P D

730 750 770
 ACGGCTCGTCAAGTTCTCCGTGCGACGCAGAGATAAGAGAGATCTATAACGAGAGTGCC
 -----+-----+-----+-----+-----+-----+-----+
 T A S S S P S T Q R *
 G L V K F S V D A E I R E I Y N E E C P

790 810 830
 CCGTCCGTCACTGACGTGTCCTCCCTCCCTCGACGGCCGCCAGTGGAGCCTCTCGATTCTTCT
 -----+-----+-----+-----+-----+-----+-----+
 V V T D V S V P L D G R Q W S L S I F S

III.3a.3

850 870 890
 CCTTCCGATGTTCAAGAACCGCCTACGTCGCCGTAGCGAACGTCGAGAACAGGAGATGT
 -----+-----+-----+-----+-----+-----+-----+
 F P M F R T A Y V A V A N V E N K E M S

910 930 950
 CGCTCGACGTTGTCAACGACCTCATCGAGTGGCTCAACAATCTGCCGACTGGCGTTATG
 -----+-----+-----+-----+-----+-----+-----+
 L D V V N D L I E W L N N L A D W R Y V

970 990 1010
 TCGTTGACTCTGAACAGTGGATTAACTTCACCAATGACACCACTACTACGTCGCCATCC
 -----+-----+-----+-----+-----+-----+-----+
 V D S E Q W I N F T N D T Y Y V R I R

1030 1050 1070
 GCGTTCTACGTCCAACCTACGACGTTCCAGACCCACAGAGGGCCTTGTTCCGACAGTCT
 -----+-----+-----+-----+-----+-----+-----+
 V L R P T Y D V P D P T E G L V R T V S

1090 1110 1130
 CAGACTACCGCCTCACTTATAAGCGGATAACATGTGAAGCCAACATGCCAACACTCGTCG
 -----+-----+-----+-----+-----+-----+-----+
 D Y R L T Y K A I T C E A N M P T L V D

III 3a.4

1450 1470 1490
ACGACGGCTCTTGACCGACTTCGCTTCGAGGAGACACCGTCACTTCCGGCAGGTCG
-----+-----+-----+-----+-----+-----+
D G S W T D F A S A G D T V T F R Q V A

1510 1530 1550
CCGTCGACGAGTCGTTGTGACCAACAACCCCGCGGCGGAGCGCCCCACCTTCA
-----+-----+-----+-----+-----+-----+
V D E V V V T N N P A G G G S A P T F T

1570 1590 1610
CCGTGAGAGTGCCCTTCAAACGCTTACACCAACACCGTGTTAGGAACACGCTCTTAG
-----+-----+-----+-----+-----+-----+
V R V P P S N A Y T N T V F R N T L L E

1630 1650 1670
AGACTCGACCTCCTCTCGTAGGCTCGAAGTCCCTATGCCACCTGCTGACTTTGGACAGA
-----+-----+-----+-----+-----+-----+
T R P S S R R L E L P M P P A D F G Q T

1690 1710 1730
CGGTCGCCAACAACCGAAGATCGAGAGTCGCTTCTTAAAGAAACACTTGGCTGCTATT
-----+-----+-----+-----+-----+-----+
V A N N P K I E Q S L L K E T L G C Y L

1750	1770	1790
TGGTCCACTCCAAATGCGAAACCCCGTTTCCAGCTCACGCCAGCCAGCTCCTTTGGCG		
-----+	-----+	-----+
V H S K M R N P V F Q L T P A S S F G A		
1810	1830	1850
CCGTTTCCCTTCAACAATCCGGGTATGAGCGCACACGGACCTCCCGGACTACACTGGCA		
-----+	-----+	-----+
V S F N N P G Y E R T R D L P D Y T G I		
1870	1890	1910
TCCGTGACTCATTCGACCAGAAACATGTCCACCGCTGTGGCCCACTTCGGCTCACTCTCCC		
-----+	-----+	-----+
R D S F D Q N M S T A V A H F R S L S H		
1930	1950	1970
ACTCCTGCAGTATCGTCACTAAGACCTACCAGGGTTGGGAAGCGTCACGAACGTCAACA		
-----+	-----+	-----+
S C S I V T K T Y Q G W E G V T N V N T		
1990	2010	2030
CGCCTTTCGGCCAATTGCGGCACGGGGCTCCTCAAGAAATGAGGAGATCCTCTGCCCTCG		
-----+	-----+	-----+
P F G Q F A H A G L L K N E E I L C L A		

370 390 410
GCAGGATGGGAGATGCTGGAGTGGCGTACACAGCGACCTCACAACCGTCGCGGAACCCGTA
-----+-----+-----+-----+-----+-----+
A G W E M L E W R H S D L T T V A E P V

430 450 470
ACGTTCCGGTCAGCGCCAACACCGTCACCGTCAATGGTAGAAGAAACCAACGGCGTCGGA
-----+-----+-----+-----+-----+-----+
T F G S A P T P S P S M V E E T N G V G

490 510 530
CCGGAAGGCAAGTTTCTCCCCCTGACAAATTTCACCGCTGCTGCACAAGACCTCGCGCAA
-----+-----+-----+-----+-----+-----+
P E G K F L P L T I S P L L H K T S R K

550 570 590
GCCTTGACGCCAACACCGTCACCTTCCCCCGCTAACATCTCTAGCATGCCCGAATTCCGG
-----+-----+-----+-----+-----+-----+
A L T P T P S L S P A N I S S M P E F R

610 630 650
AATGGGCCAAGGAAAGATCGACCTCGACTCCGATTCCATCGGCTGGTACTTCAAGTAC
-----+-----+-----+-----+-----+-----+
N W A K G K I D L D S D S I G W Y F K Y

III 3b2

670 690 710
 CTTGACCCAGCGGTGCTACAGAGTCTGCGCGCCGTCGGGAGTACTCGAAGATCCCT
 -----+-----+-----+-----+-----+-----+
 L D P A G A T E S A R A V G E Y S K I P

730 750 770
 GACGGCCTCGTCAAGTTCCTCGTCGACGCAGAGATAAGAGAGATCTATAACGAGGAGTGC
 -----+-----+-----+-----+-----+-----+
 D G L V K F S V D A E I R E I Y N E E C

790 810 830
 CCCGTCGTCACGTGTCGTCCTCCCTCGACGGCCGACGTGGAGCCTCTCGATTTC
 -----+-----+-----+-----+-----+-----+
 P V V T D V S V P L D G R Q W S L S I F

850 870 890
 TCCTTCCGATGTTCAAGAACCGCCTACGTCCGCTAGCGAACGTCGAGAACAGGAGATG
 -----+-----+-----+-----+-----+-----+
 S F P M F R T A Y V A V A N V E N K E M

910 930 950
 TCGCTCGACGTTGTCAACGACCTCATCGAGTGGCTCAACAATCTCGCCGACTGGCGTTAT
 -----+-----+-----+-----+-----+-----+
 S L D V V N D L I E W L N N L A D W R Y

III 36.3

970 990 1010
 GTCGTTGACTCTGAACAGTGGATTAACTTCACCAATGACACACGTAAGTACTGTCGCGCATC
 -----+-----+-----+-----+-----+-----+-----+
 V V D S E Q W I N F T N D T T Y Y Y V R I

1030 1050 1070
 CGCGTTCTACGTCCAACCTACGACGTTCCAGACCCACAGAGGGCCTTGTTCGCACAGTC
 -----+-----+-----+-----+-----+-----+-----+
 R V L R P T Y D V P D P T E G L V R T V

1090 1110 1130
 TCAGACTACCGCCTCACTTATAAGGCGATAACATGTGAAGCCAACATGCCAACACTCGTC
 -----+-----+-----+-----+-----+-----+-----+
 S D Y R L T Y K A I T C E A N M P T L V

1150 1170 1190
 GACCAAGGCTTTGGATCGGCGGCCAGTACGCTCTCACCCTCAGTACTAGCCTACCGCAGTAC
 -----+-----+-----+-----+-----+-----+-----+
 D Q G F W I G G Q Y A L T P T S L P Q Y

1210 1230 1250
 GACGTCAGCGAGGCCCTACGCTCTGCACACTTTGACCTTCGCCAGACCATCCAGCGCGCT
 -----+-----+-----+-----+-----+-----+-----+
 D V S E A Y A L H T L T F A R P S S A A

III 3b.4

1270 1290 1310
GCACTCGCGTTTGTGTGGCAGGTTTGCCACAGGGTGGCAGTGCCTGCAGGCACTCCA
-----+-----+-----+-----+-----+-----+
A L A F V W A G L P Q G G T A P A G T P

1330 1350 1370
GCCTGGGAGCAGGCATCCTCGGGTGGCTACCTCAGCTGGCGCCACAACGGTACTTTC
-----+-----+-----+-----+-----+-----+
A W E Q A S S G G Y L T W R H N G T T F

1390 1410 1430
CCAGCTGGCTCCGTTAGCTACGTTCTCCTGAGGGTTTCGCCCTTGAGCGCTACGACCCG
-----+-----+-----+-----+-----+-----+
P A G S V S Y V L P E G F A L E R Y D P

1450 1470 1490
AACGACGGCTCTTGACCGACTTCGCTTCGCGAGGAGACACCGTCACTTTCGGCAGGTC
-----+-----+-----+-----+-----+-----+
N D G S W T D F A S A G D T V T F R Q V

1510 1530 1550
GCCGTCGACGAGGTCGTTGTGACCAACAACCCCGCGGGCGGCGAGCGCCCCACCTTC
-----+-----+-----+-----+-----+-----+
A V D E V V V T N N P A G G S A P T F

III-3b.5

1570 1590 1610
ACCGTGAGAGTGCCCCCTTCAACGCTTACACCAACCGTGTTAGGAACACGCTCTTA
-----+-----+-----+-----+-----+-----+
T V R V P P S N A Y T N T V F R N T L L

1630 1650 1670
GAGACTCGACCTCCTCTCGTAGGCTCGAACTCCCTATGCCACCTGCTGACTTTGGACAG
-----+-----+-----+-----+-----+-----+
E T R P S S R R L E L P M P P A D F G Q

1690 1710 1730
ACGGTCGCCAACAACCCGAAGATCGAGCAGTCGCTTCTTAAAGAAACACTTGGCTGCTAT
-----+-----+-----+-----+-----+-----+
T V A N N P K I E Q S L L K E T L G C Y

1750 1770 1790
TTGGTCCACTCCAATAATGCGAAACCCCGTTTTCAGCTCACGCCAGCCAGCTCCTTTGGC
-----+-----+-----+-----+-----+-----+
L V H S K M R N P V F Q L T P A S S F G

1810 1830 1850
GCCGTTTCCTTCAACAATCCGGGTATGAGCGCACACGCGACCTCCCGGACTACACTGGC
-----+-----+-----+-----+-----+-----+
A V S F N N P G Y E R T R D L P D Y T G

III 3b.6

1870 1890 1910
ATCCGTGACTCATTCGACCAGAACATGTCCACCGCTGTGGCCCACTTCCGGCTCACTCTCC
-----+-----+-----+
I R D S F D Q N M S T A V A H F R S L S

1930 1950 1970
CACTCCTGCAGTATCGTCACTAAGACCTACCAGGGTTGGAGGCGTCAAGAACGTCAAC
-----+-----+-----+
H S C S I V T K T Y Q G W E G V T N V N

1990 2010 2030
ACGCCTTTCGGCCAAATTCGGGCACGGGGCCCTCCTCAAGAATGAGGAGATCCTCTGCCTC
-----+-----+-----+
T P F G Q F A H A G L L K N E E I L C L

2050 2070 2090
GCCGACGACCTGGCCACCCGTCTCACAGGTGTCTACCCCGCCCACTGACAACTTCGGGGCC
-----+-----+-----+
A D D L A T R L T G V Y P A T D N F A A

2110 2130 2150
GCCGTTTCTGCTTCGGCCGGAACATGTGTCTCCTCGTCTGAAGTCGGAGGCAACGTCC
-----+-----+-----+
A V S A F A A N M L S S V L K S E A T S

2170 2190 2210
TCCATCATCAAGTCCGTTGGCGAGACTGCCGTCGGCGGCTCAGTCCGGCCTCGCGAAG
-----+-----+-----+-----+-----+
S I I K S V G E T A V G A A Q S G L A K

2230 2250 2270
CTACCCGACTGCTAATGAGTGTAACAGGGAAGATTGCCGCGCTGTCTCCGGCGGCCGA
-----+-----+-----+-----+-----+
L P G L L M S V P G K I A A R V R A R R

2290 2310 2330
GCGCGCCGCGCGCTCGTGCCAAATTAGTTTGCTCGCTCCTGTTTCGCCGTTTCGTAA
-----+-----+-----+-----+-----+
A R R R A A R A N *

2350 2370 2390
AACGGCGTGGTCCCGCACATTACGGGTACCCCTAAAGACTCTGGTGAGTCCCCCGTGTAC
-----+-----+-----+-----+-----+
-----+-----+-----+-----+-----+

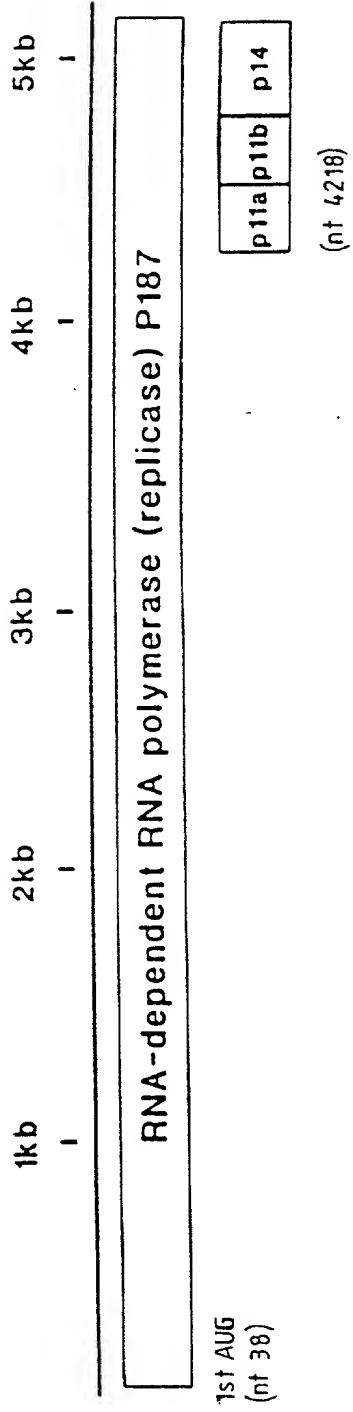
2410 2430 2450
ACGACGGGTCTGCCGCGGTTCGATTCCATTCCCAAGCGGCAAGACGTAGTAGCTC
-----+-----+-----+-----+-----+
-----+-----+-----+-----+-----+

2470
TGCGTCCCTCGGGATACCA
-----+-----+-----+-----+-----+

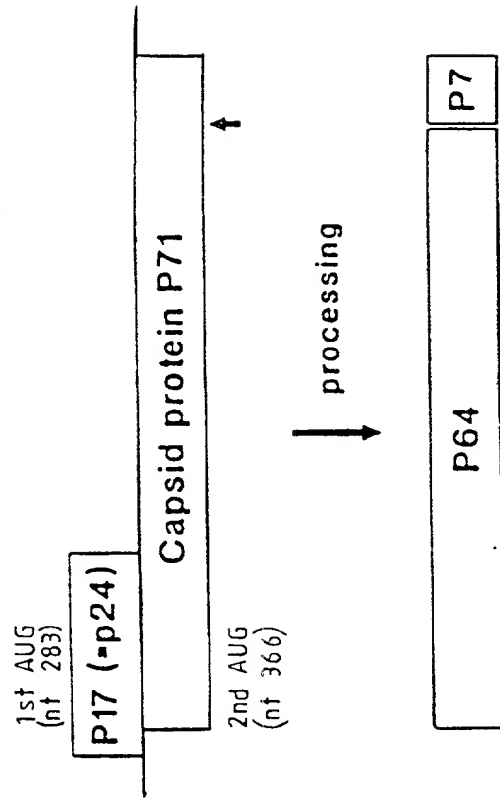
III 3b.8

Proteins encoded by the HaSV genome

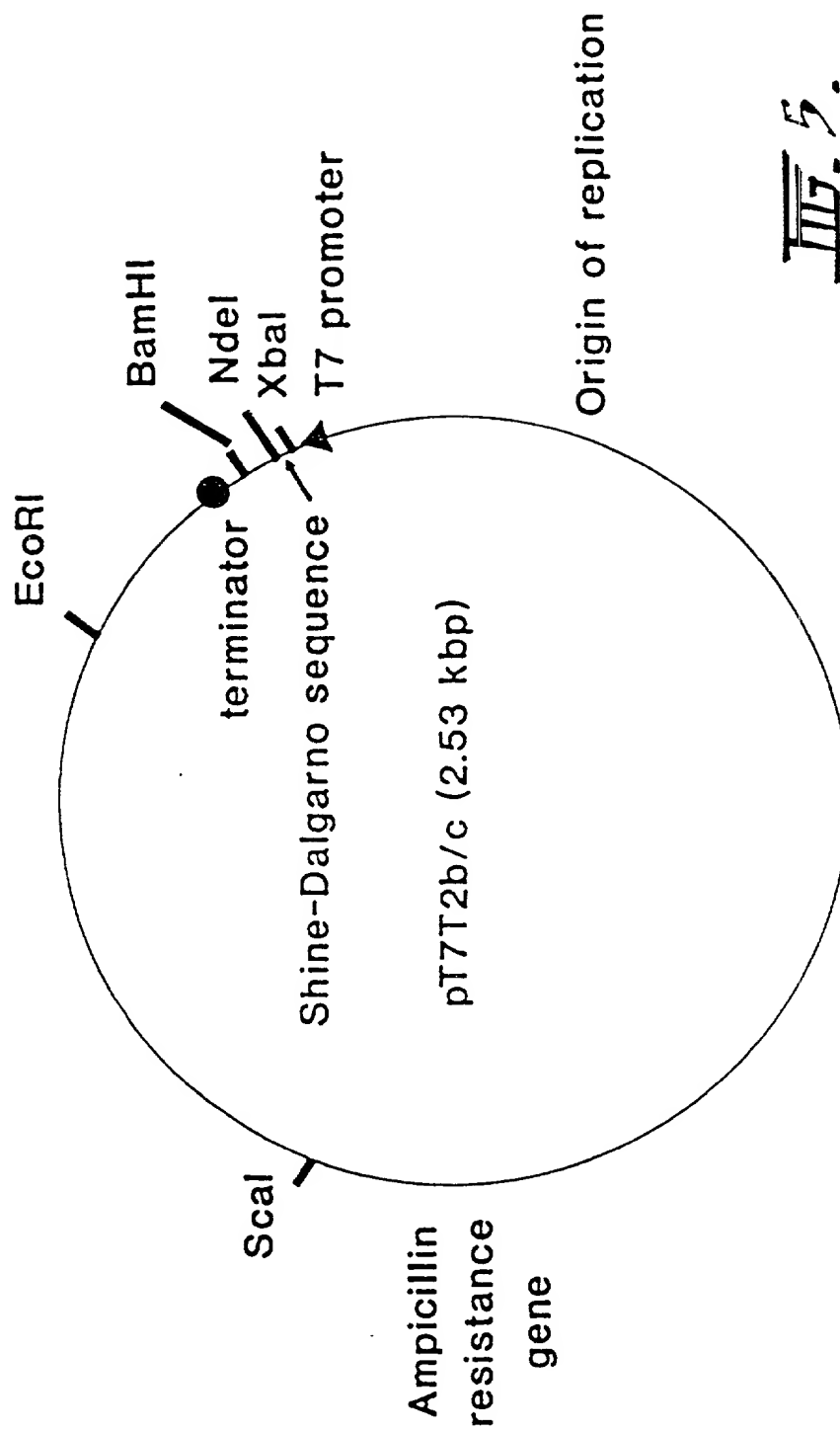
RNA 1



RNA 2

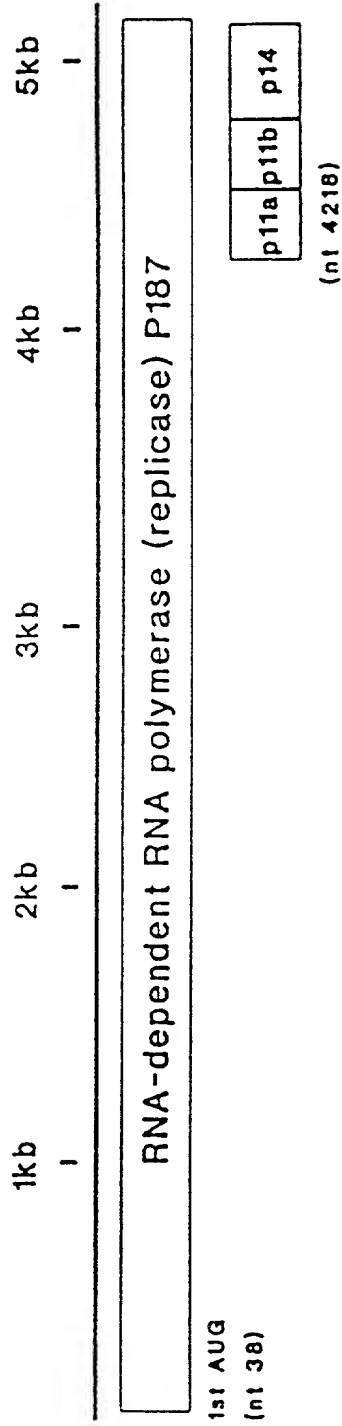


III. 4.

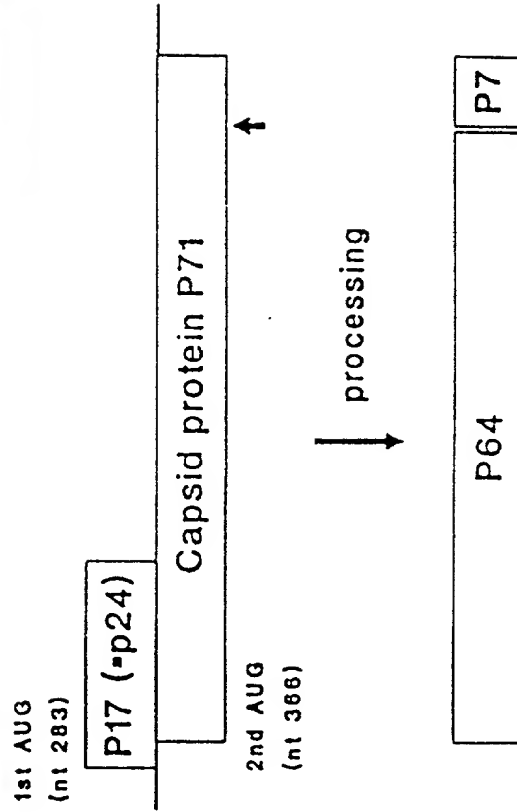
III.5.

Proteins encoded by the HaSV genome

RNA 1



RNA 2



PROTEINS EXPRESSED FROM HaSV RNA 2

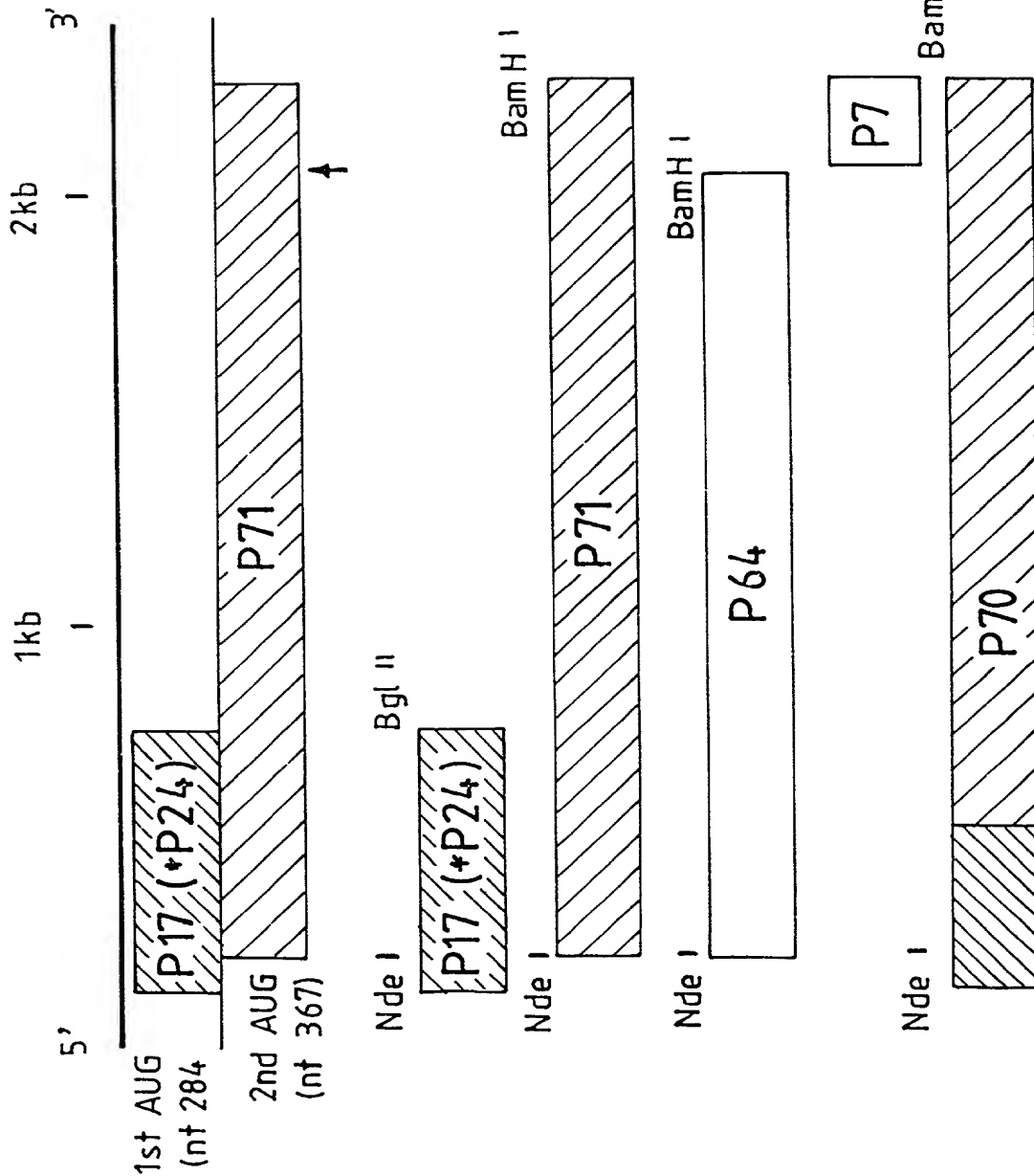
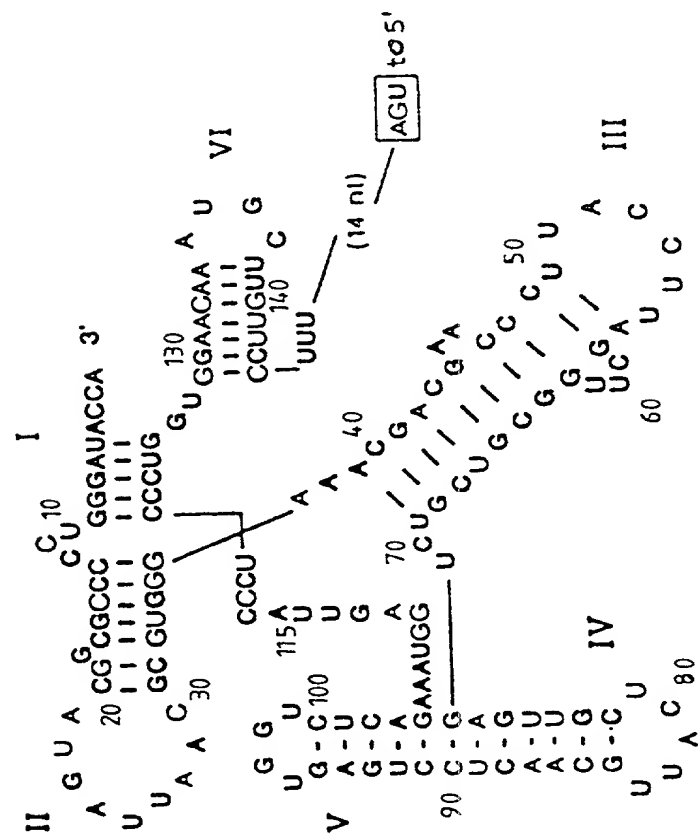


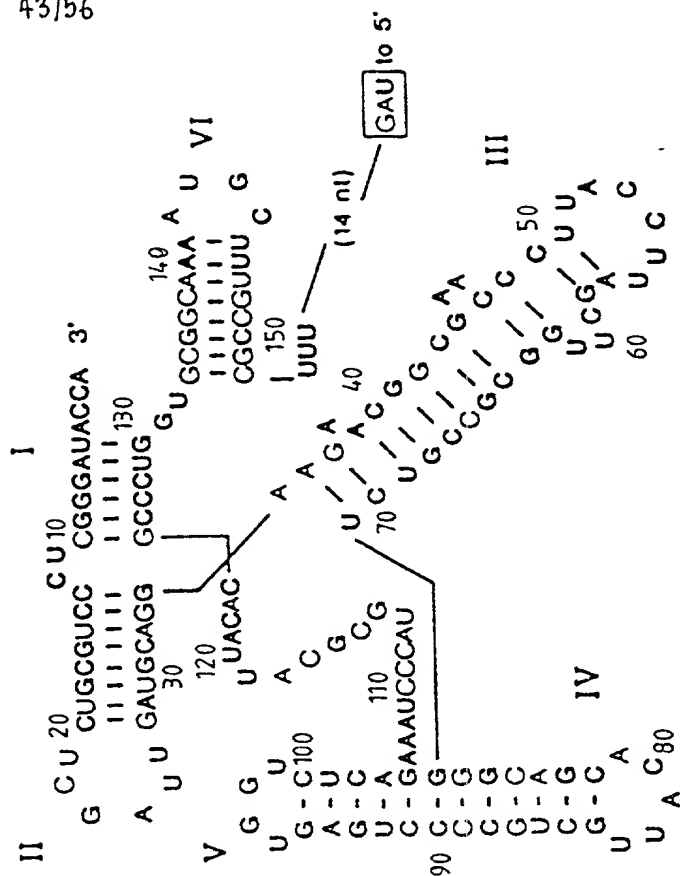
Fig. 7.

HaSV RNA 3' - terminal tRNA-like structures

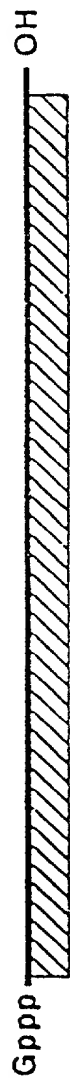
RNA 1



RNA 2



HaSV RNA1



HaSV RNA2



Insect cell expression constructs

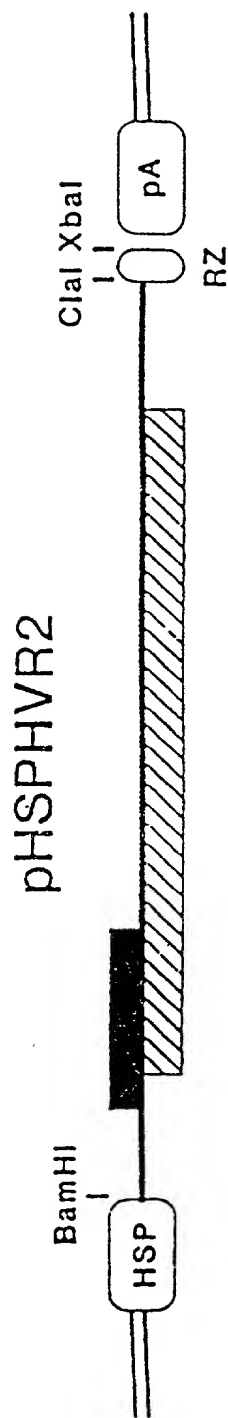
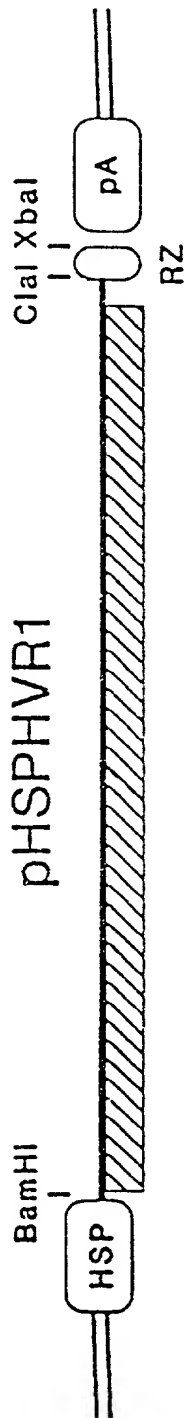
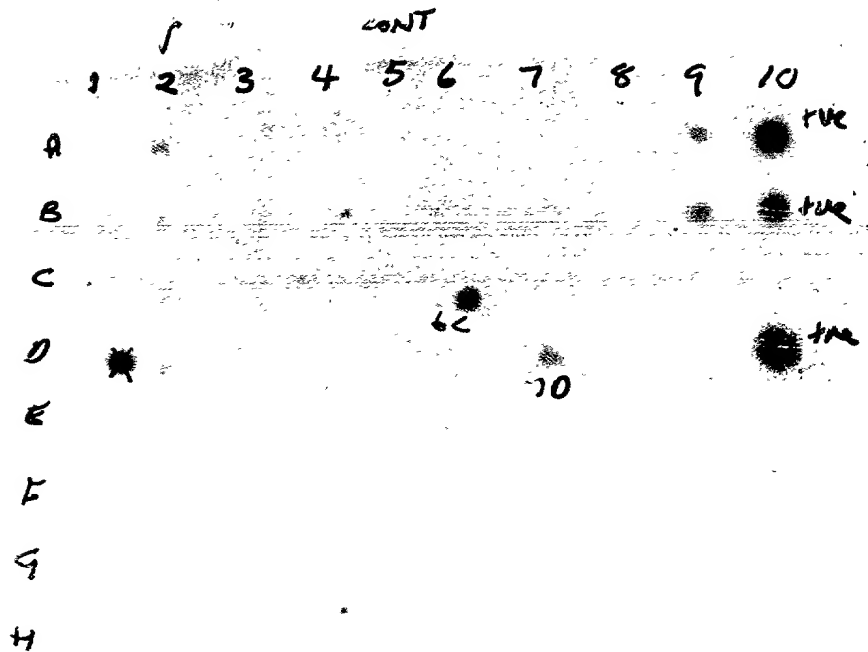


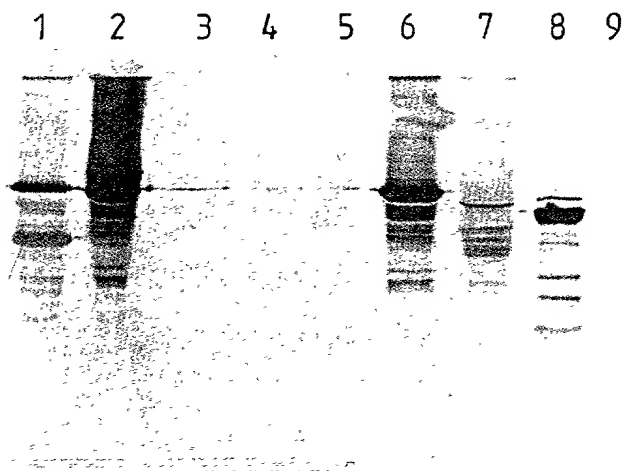
Figure 1 consists of seven sub-graphs, labeled (a) through (g), each showing the growth of *E. coli* O157:H7 in ground beef over a 24-hour period under different treatment conditions. The y-axis for all graphs is \log_{10} CFU/g, ranging from 0 to 8. The x-axis is time in hours, ranging from 0 to 24. Graph (a) is the control, showing a steady increase in bacterial count from approximately 1.5 to 7.5. Graphs (b) through (g) show the effect of 100% NaCl, 100% NaOH, 100% H_2O_2 , 100% HCl, 100% acetic acid, and 100% lactic acid, respectively. In all treated samples, the bacterial count is significantly lower than in the control, with some treatments (like NaOH and H_2O_2) showing a more pronounced and sustained reduction.



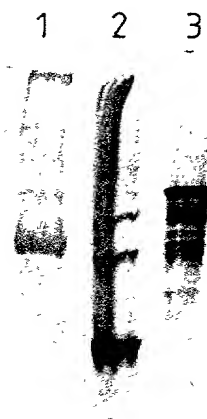
III. 11.

WESTERN BLOTS OF HaSV CAPSID PROTEIN

A. HaSV ANTISERUM



B. HaSV ANTISERUM



C. Bt ANTISERUM

1 2 3

Bt toxin <ts

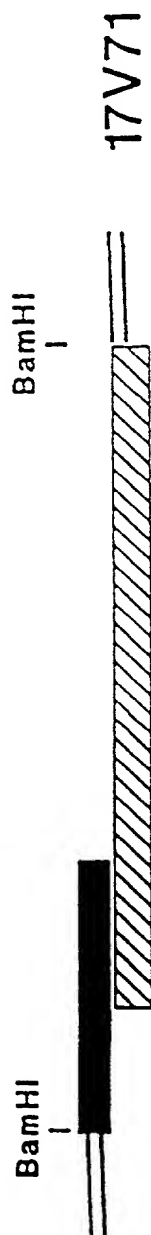
AUG context



P17: UUaugA
P71: AGGaugG

Baculovirus Constructs

III. 13.



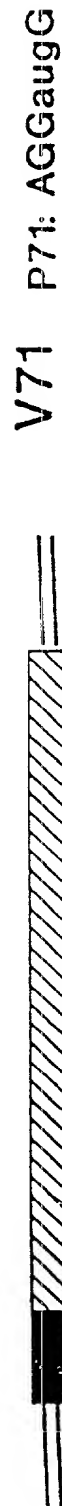
17V71
P17: UUUangA
P71: AGGangG



17E71
P17: GACaugA
P71: AGGaugG



P71 P71: GACa₉A



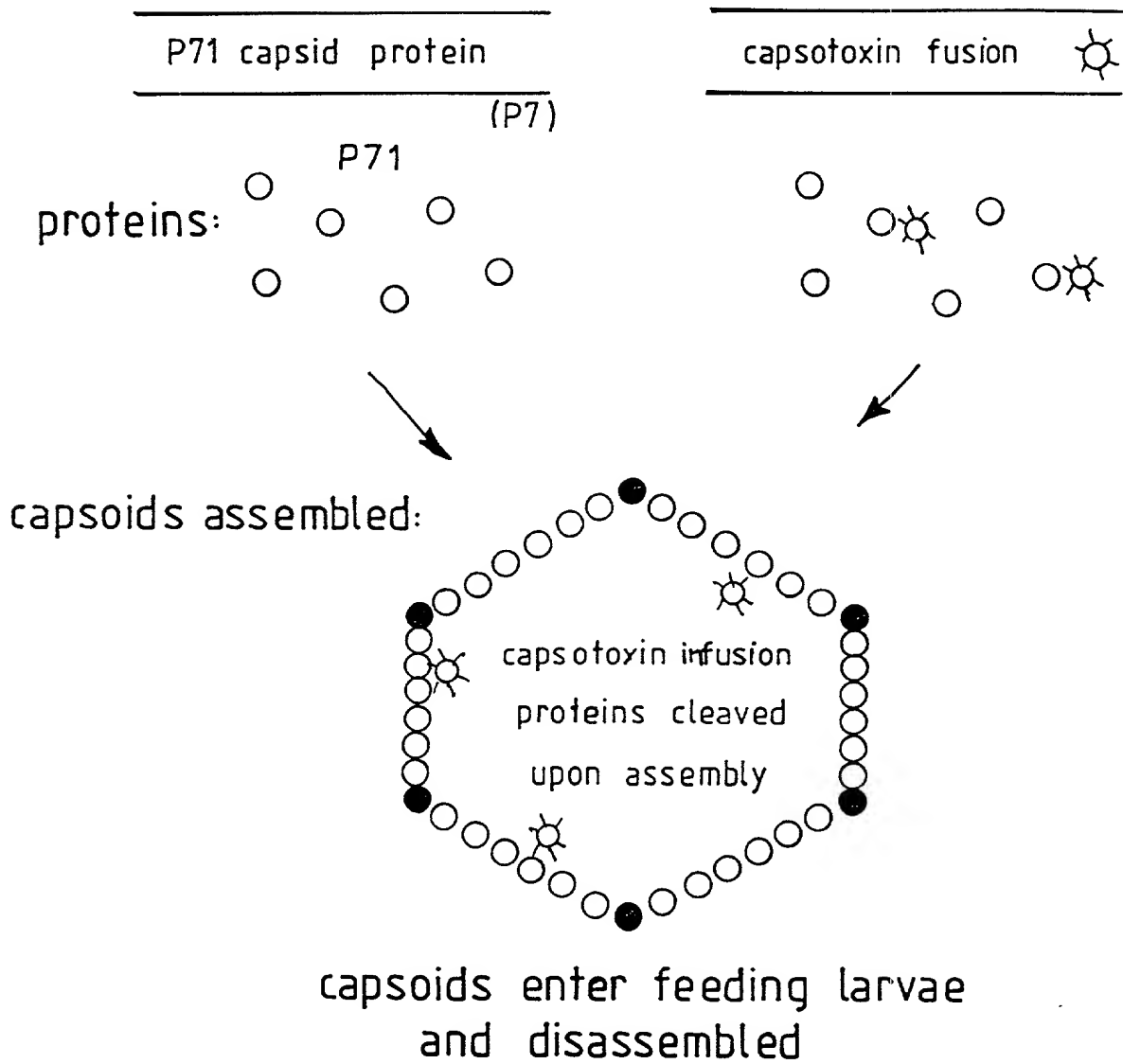
V71 P71: AGGaugG



R2 P17: UUUaugA
P71: AGGaugG

Virus capsid strategy: capsotoxin encapsulation

transgenic plant genome:



releasing active insect-specific toxin

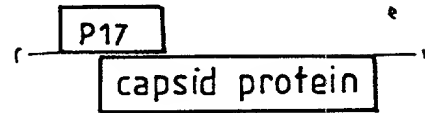
Virus capsoid strategy:

toxin message encapsulation and amplification

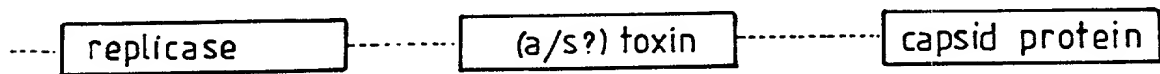
HaSV RNA 1



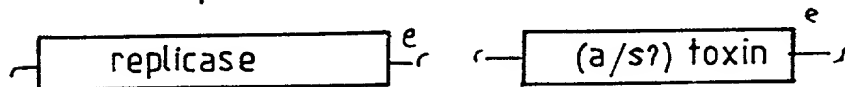
HaSV RNA 2



transgenic plant genome:

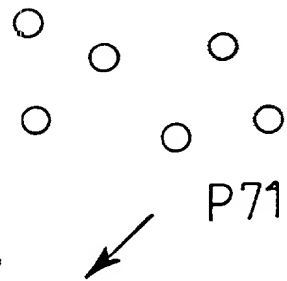
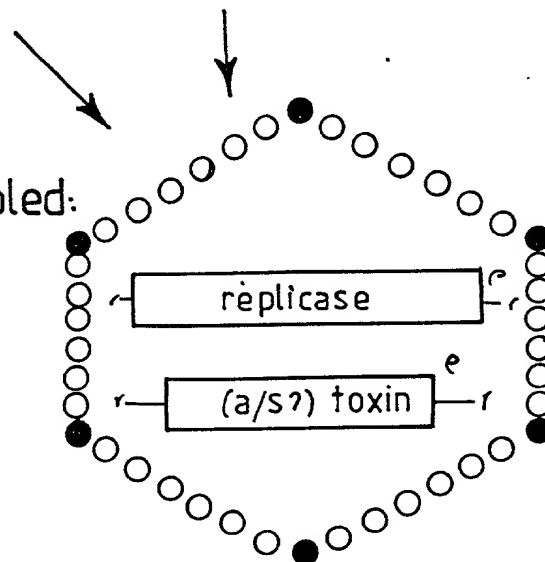


transcripts:



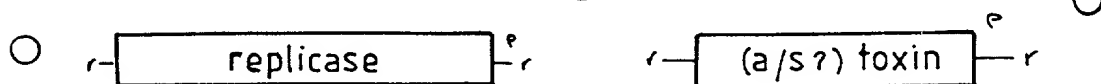
proteins:

capsoids assembled:



capsoids enter and infect feeding larvae

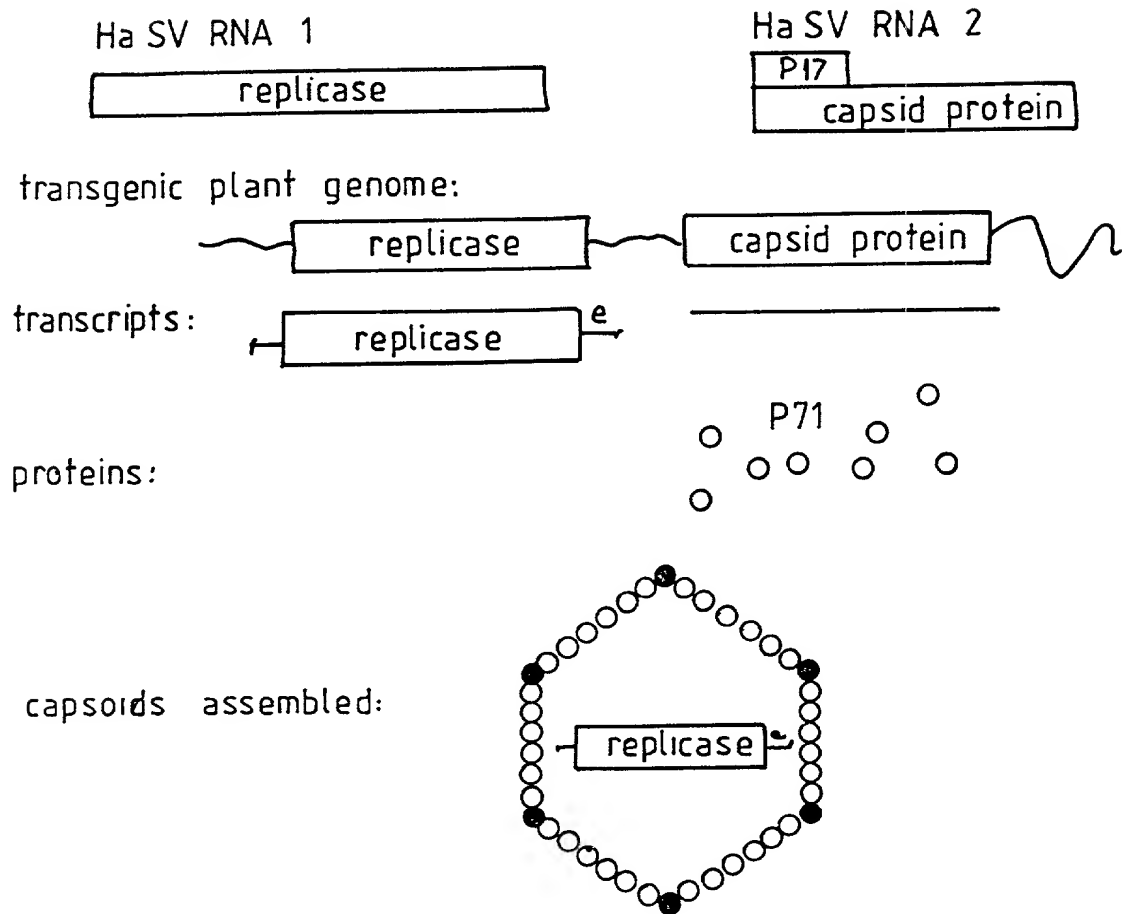
capsoids disassembled:



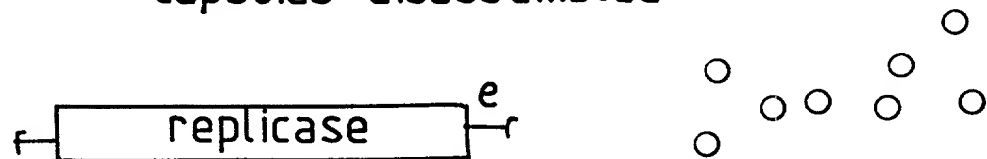
mRNA amplification and expression and secretion of toxin

Fig. 14b

Virus expression in plants: the one-way vector



capsoids enter and infect feeding larvae
capsoids disassembled:



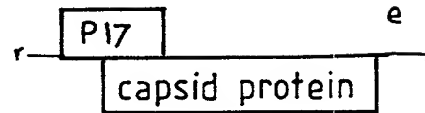
RNA 1 replicates and causes
anti-feeding effect:

Fig. 14c

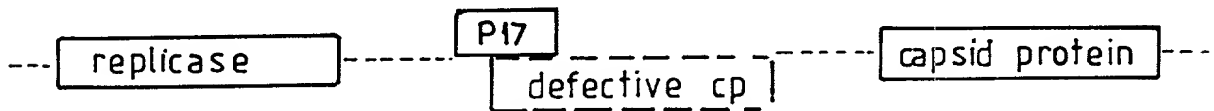
HaSV RNA1



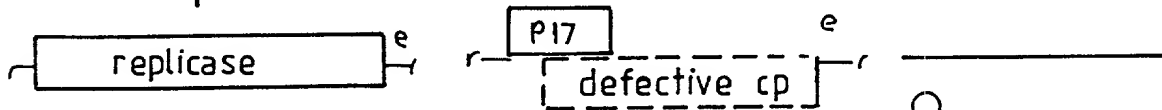
HaSV RNA



transgenic plant genome:

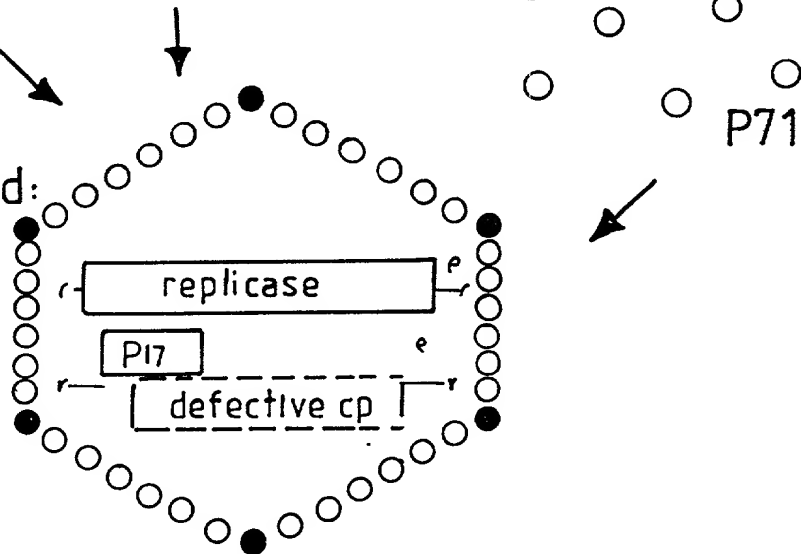


transcripts:



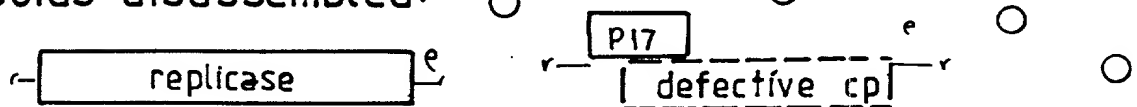
proteins:

capsoids assembled:



capsoids enter and infect feeding larvae

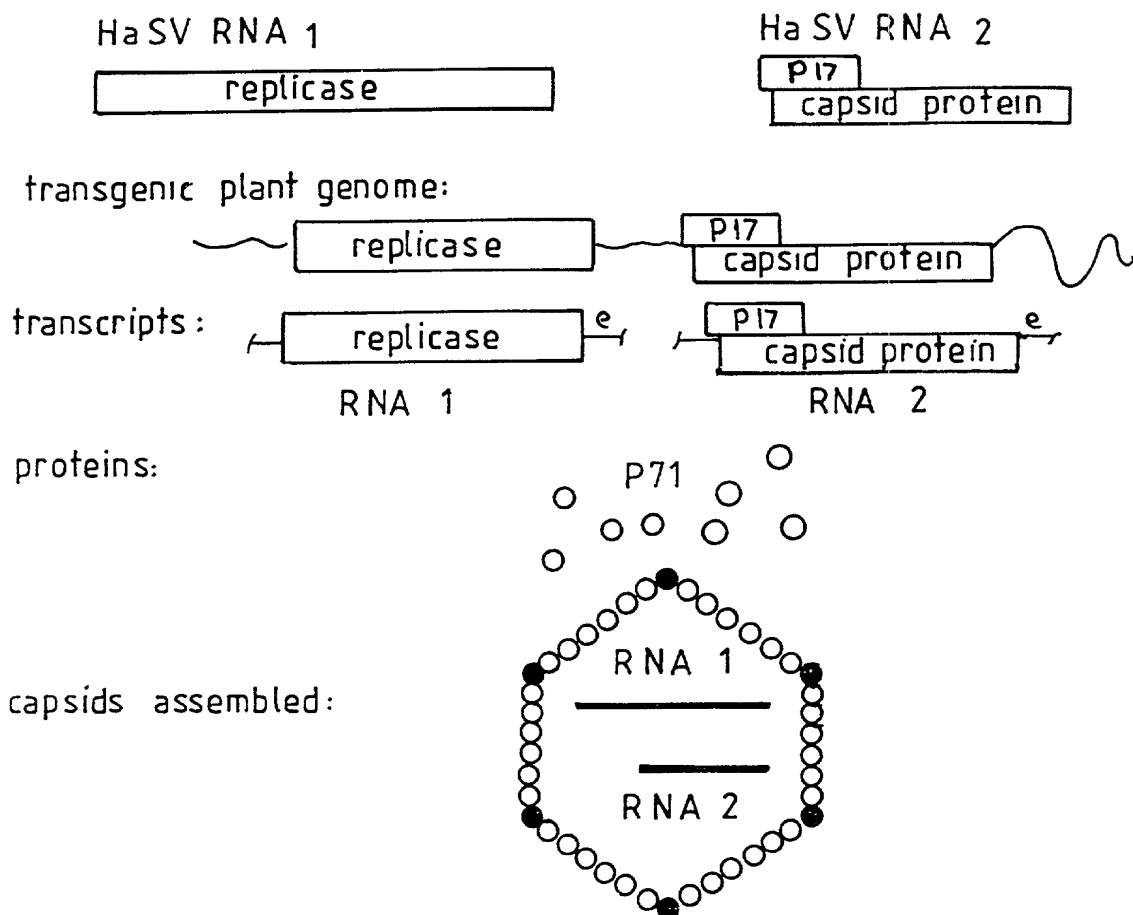
capsoids disassembled:



viral RNA replicates and causes anti-feeding effect

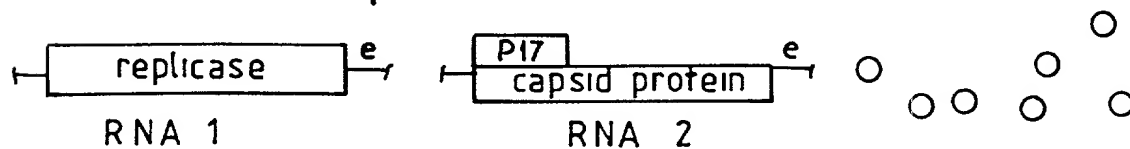
III. 14d.

Virus expression in plants



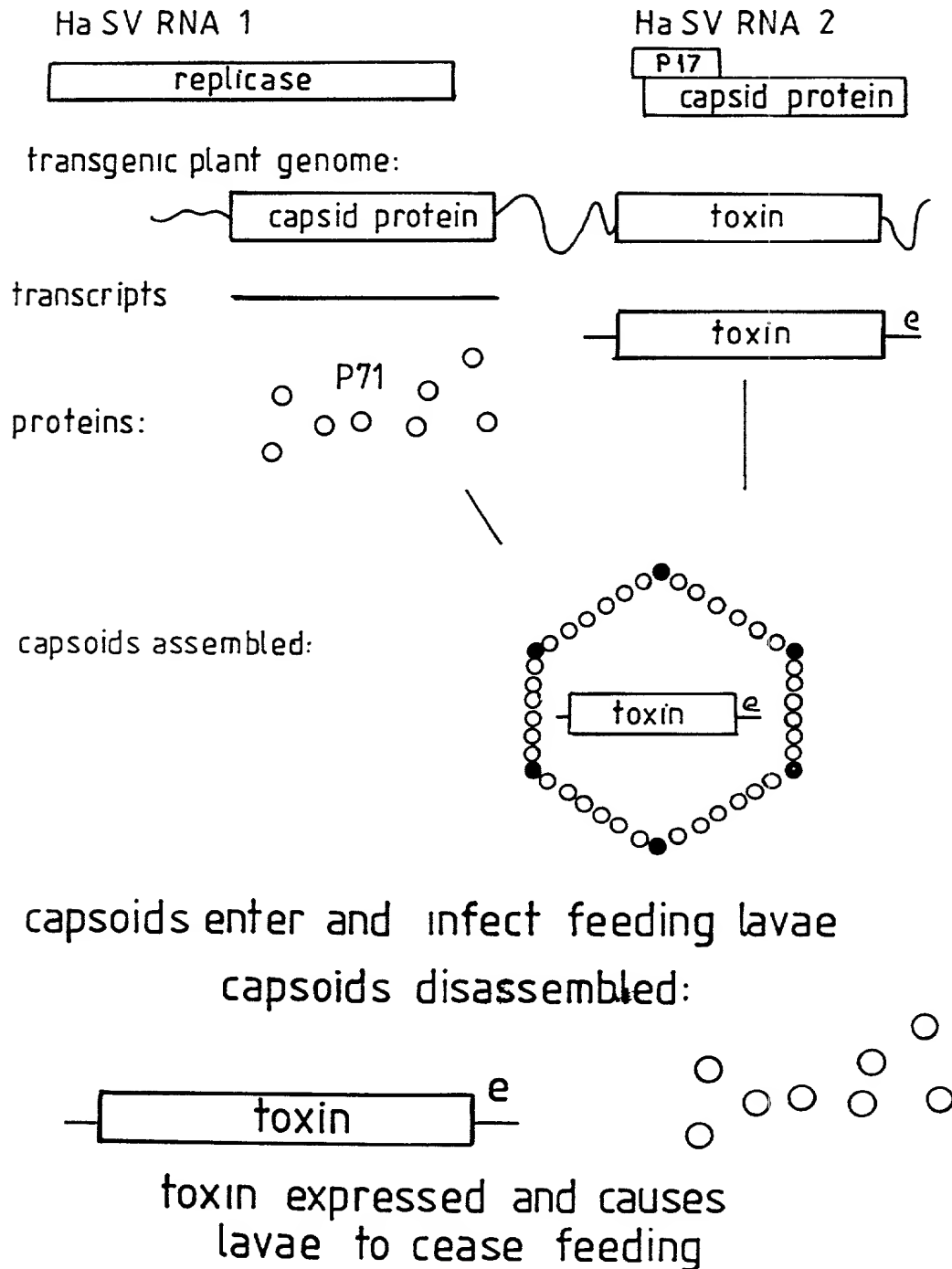
capsids enter and infect feeding larvae

capsids disassembled:



virus replicates and causes
anti-feeding effect

Virus expression in plants: the one-way vector for a toxin



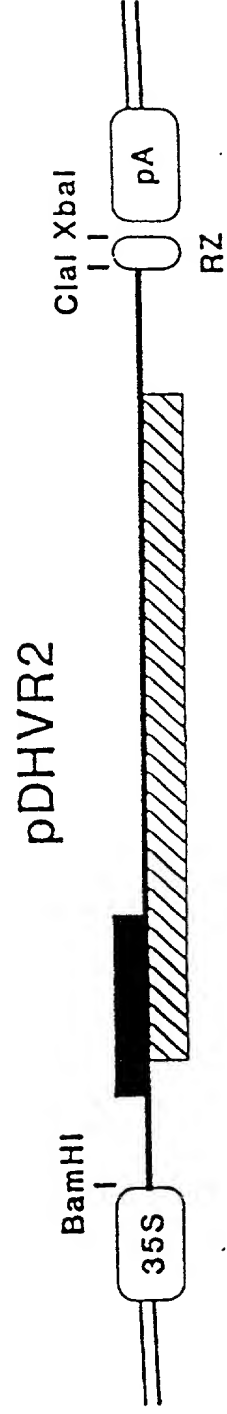
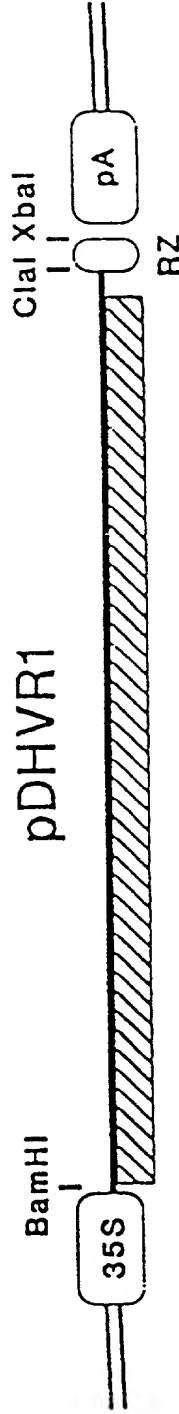
HaSV RNA1



HaSV RNA2



Protoplast expression constructs



III 16.

HaSV RNA1

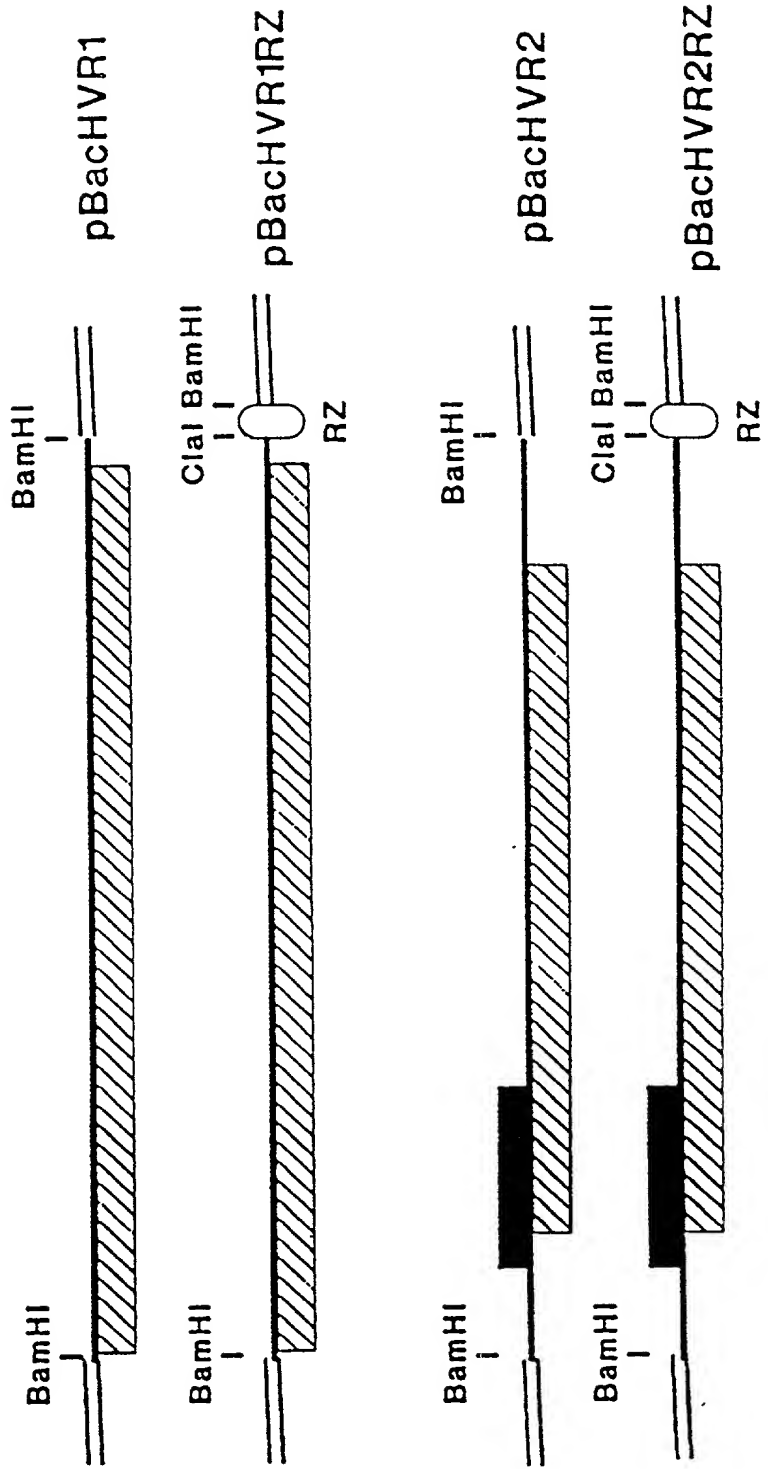


HaSV RNA2



III.15.

baculovirus expression constructs



DECLARATION AND POWER OF ATTORNEY
FOR PATENT APPLICATION

As a below-named inventor, I hereby declare that:

My residence, post office address and citizenship are as stated below next to my name,

I believe I am the original, first and sole inventor (if only one name is listed below) or an original, first and joint inventor (if plural names are listed below) of the subject matter which is claimed and for which a patent is sought on the invention entitled INSECT VIRUSES AND THEIR USES IN PROTECTING PLANTS

the specification of which

(check ☐ is attached hereto.
one)

☒ was filed on June 7, 1995 as
Application Serial No. 08/485,355
and was amended on _____
(if applicable)

I hereby state that I have reviewed and understand the contents of the above-identified specification, including the claims, as amended by any amendment referred to above.

I acknowledge the duty to disclose to the Patent Office all information known to me to be material to patentability as defined in 37 C.F.R. 1.56.

I hereby claim foreign priority benefits under Title 35, United States Code, §119 of any foreign application(s) for patent or inventor's certificate listed below and have also identified below any foreign application for patent or inventor's certificate having a filing date before that of the application on which priority is claimed:

Prior Foreign Application(s)			Priority Claimed	
<u>PL4081/92</u>	<u>Australia</u>	<u>14 August 1992</u>	<input checked="" type="checkbox"/>	<input type="checkbox"/>
(Number)	(Country)	(Day/Month/Year Filed)	Yes	No
_____	_____	_____	<input type="checkbox"/>	<input type="checkbox"/>
(Number)	(Country)	(Day/Month/Year Filed)	Yes	No
_____	_____	_____	<input type="checkbox"/>	<input type="checkbox"/>
(Number)	(Country)	(Day/Month/Year Filed)	Yes	No

I hereby claim the benefit under Title 35, United States Code, §120 of any United States application(s) listed below and, insofar as the subject matter of each of the claims of this application is not disclosed in the prior United States application in the manner provided by the first paragraph of Title 35, United States Code, §112, I acknowledge the duty to disclose to the Patent Office all information known to me to be material to patentability as defined in 37 C.F.R. 1.56 which occurred between the filing date of the prior application and the national or PCT international filing date of this application:

<u>08/440,522</u>	<u>12 May 1995</u>	<u>Pending</u>
(Application Serial No.)	(Filing Date)	(Status)
		(patented, pending, abandoned)
_____	_____	_____
(Application Serial No.)	(Filing Date)	(Status)
		(patented, pending, abandoned)

I hereby appoint the following attorneys to prosecute this application and to transact all business in the Patent and Trademark Office connected therewith: Harold C. Hohbach, Reg. No. 17,757; Aldo J. Test, Reg. No. 18,048; Thomas O. Herbert, Reg. No. 18,612; Donald N. MacIntosh, Reg. No. 20,316; Jerry G. Wright, Reg. No. 20,165; Edward S. Wright, Reg. No. 24,903; David J. Brezner, Reg. No. 24,774; Richard E. Backus, Reg. No. 22,701; James A. Sheridan, Reg. No. 25,435; Robert B. Chickering, Reg. No. 24,286; Gary S. Williams, Reg. No. 31,066; Richard F. Trecartin, Reg. No. 31,801; C. Michael Zimmerman, Reg. No. 20,451; Walter H. Dreger, Reg. No. 24,190; Steven F. Caserza, Reg. No. 29,780; _____;

provided that if any one of said attorneys ceases being affiliated with the law firm of Flehr, Hohbach, Test, Albritton & Herbert as partner, employee or of counsel, such attorney's appointment as attorney and all powers derived therefrom shall terminate on the date such attorney ceases being so affiliated.

Direct all telephone calls to Walter H. Dreger at (415) 781-1989.

Address all correspondence to:

FLEHR, HOHBACH, TEST,
ALBRITTON & HERBERT
Suite 3400, Four Embarcadero Center
San Francisco, California 94111

File No. A-58631-2/WH/D/LEA

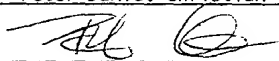
I hereby declare that all statements made herein of my own knowledge are true and that all statements made on information and belief are believed to be true; and further that these statements were made with the knowledge that willful false statements and the like so made are punishable by fine or imprisonment, or both, under Title 18, United States Code, §1001 and that such willful false statements may jeopardize the validity of the application or any patent issued thereon.

Full name of sole or

first inventor:

Peter Daniel Christian

Inventor's signature:



Date:

23/6/95

Residence:

MAGUIRE

Lynham, Australian Capital Territory, 2601 Australia

Citizenship:

Great Britain

Post Office Address:

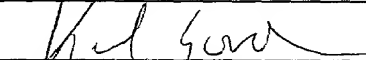
8a Corin Court, Wattle Street PO Box 13 OLIVIND

Lynham, Australian Capital Territory, 2601 Australia

Full name of second joint
inventor, if any:

Karl Heinrich Julius Gordon

Inventor's signature:



Date:

22 June 1995

Residence:

Weston, Australian Capital Territory, 2611 Australia

Citizenship:

Australia/Germany

Post Office Address:

18 Chevalier Street

Weston, Australian Capital Territory, 2611 Australia

